

Genetic Diversity of the Q Fever Agent, *Coxiella burnetii*, Assessed by Microarray-Based Whole-Genome Comparisons†

Paul A. Beare,¹ James E. Samuel,² Dale Howe,¹ Kimmo Virtaneva,³
Stephen F. Porcella,⁴ and Robert A. Heinzen^{1*}

Coxiella Pathogenesis Section, Laboratory of Intracellular Parasites,¹ Laboratory of Persistent Viral Diseases,³ and Genomics Core Facility,⁴ Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, Montana 59840, and Department of Medical Microbiology and Immunology, Texas A&M University System, Health Science Center, College Station, Texas 77843²

Received 19 September 2005/Accepted 15 December 2005

Coxiella burnetii, a gram-negative obligate intracellular bacterium, causes human Q fever and is considered a potential agent of bioterrorism. Distinct genomic groups of *C. burnetii* are revealed by restriction fragment-length polymorphisms (RFLP). Here we comprehensively define the genetic diversity of *C. burnetii* by hybridizing the genomes of 20 RFLP-grouped and four ungrouped isolates from disparate sources to a high-density custom Affymetrix GeneChip containing all open reading frames (ORFs) of the Nine Mile phase I (NMI) reference isolate. We confirmed the relatedness of RFLP-grouped isolates and showed that two ungrouped isolates represent distinct genomic groups. Isolates contained up to 20 genomic polymorphisms consisting of 1 to 18 ORFs each. These were mostly complete ORF deletions, although partial deletions, point mutations, and insertions were also identified. A total of 139 chromosomal and plasmid ORFs were polymorphic among all *C. burnetii* isolates, representing ca. 7% of the NMI coding capacity. Approximately 67% of all deleted ORFs were hypothetical, while 9% were annotated in NMI as nonfunctional (e.g., frameshifted). The remaining deleted ORFs were associated with diverse cellular functions. The only deletions associated with isogenic NMI variants of attenuated virulence were previously described large deletions containing genes involved in lipopolysaccharide (LPS) biosynthesis, suggesting that these polymorphisms alone are responsible for the lower virulence of these variants. Interestingly, a variant of the Australia QD isolate producing truncated LPS had no detectable deletions, indicating LPS truncation can occur via small genetic changes. Our results provide new insight into the genetic diversity and virulence potential of *Coxiella* species.

Coxiella burnetii, a gram-negative obligate intracellular bacterium, is the etiological agent of human Q fever. Q fever can manifest as an acute or chronic illness. Acute disease is typically a self-limiting febrile illness during which pneumonia or hepatitis can occur (41). Chronic disease, although rare, is a severe illness that usually manifests as endocarditis and occasionally as vascular infection, osteomyelitis, or chronic hepatitis (51). Infections with this zoonotic pathogen occur worldwide with the possible exception of New Zealand (41). The range of *C. burnetii* reservoirs is extensive and includes both wild and domestic mammals, birds, and arthropods such as ticks (41). In most animals, *C. burnetii* does not cause overt disease. Exceptions are sheep and goats, where massive proliferation of the organism in the female reproductive system can result in late-term abortion (41). Consequently, the results of parturition can deposit tremendous numbers of *C. burnetii* in the environment, where the agent can persist for years. Inhalation of aerosols containing *C. burnetii* is the typical route of human infection. Indeed, Q fever is considered an occupational hazard associated with domestic livestock operations.

Moreover, aerosol transmission, environmental stability and a very low infectious dose (43) make *C. burnetii* a potential biowarfare agent (USA CDC, category B).

Isolates of *C. burnetii* derived from a variety of geographical areas and various hosts display considerable genetic homogeneity when examined solely by 16S rRNA gene sequencing (65). However, restriction fragment length polymorphism (RFLP) analysis of genomic DNA (gDNA) reveals considerable heterogeneity in banding patterns (31, 33, 36, 40, 77), indicating genetic diversity between *C. burnetii* isolates. Using this method, Hendrix et al. (33) categorized 32 *C. burnetii* isolates into six distinct genomic groups (I to VI). Additional genomic variance was subsequently described by Jager et al. (36) in their RFLP analysis of 80 isolates. Differentiation of *C. burnetii* has also been achieved by sequence and/or PCR-RFLP analysis of *icd* (encoding isocitrate dehydrogenase) (5, 46), *comI* (encoding an outer membrane-associated immunoreactive protein) (34, 62, 83), and *mucZ* (known to confer a mucoid property to bacteria) (62). The most recent and extensive survey of *C. burnetii* genetic diversity was reported by Glazunova et al. (24), who used multispacer sequence typing to place 173 *C. burnetii* isolates into 30 different genotypic groups.

Most *C. burnetii* isolates harbor 1 of 4 autonomously replicating plasmids termed QpH1, QpRS, QpDV, and QpDG (40, 75). These plasmids range from 32 to 42 kb in size and share a common 25-kb “core” region along with unique regions (40, 59, 79). QpRS-like plasmid sequences are integrated into the chromosome of some isolates (40, 60, 79). Plasmid types are

* Corresponding author. Mailing address: Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, 903 S. 4th. St., Hamilton, MT 59840. Phone: (406) 375-9695. Fax: (406) 375-9380. E-mail: rheinzen@niaid.nih.gov.

† Supplemental material for this article may be found at <http://jb.asm.org/>.

associated with specific genomic groups. Interestingly, human isolates within genomic groups I, II, and III were all derived from acute disease patients, whereas human isolates within genomic groups IV and V were all derived from chronic disease patients. These correlations led to the controversial hypothesis that plasmids and their associated genome encode pathotype-specific virulence factors (36, 40, 71). Controversy aside, the absolute conservation of chromosomally integrated or autonomously replicating plasmid sequences in all *C. burnetii* isolates examined to date suggests that these sequences play a critical role in some aspect of *C. burnetii* biology.

The genetic diversity of *C. burnetii* is also revealed by production of antigenically and structurally unique lipopolysaccharide (LPS) molecules. Three distinct *C. burnetii* LPS chemotypes have been described that are associated with specific genomic groups (27, 28, 30, 73), and a potential link between LPS chemotype and *C. burnetii* virulence potential has been proposed (27). LPS is the only defined virulence factor of *C. burnetii* (43). Virulent *C. burnetii* isolated from natural sources and infections all produce a full-length LPS that is serologically defined as "phase I." Serial in vitro passage of phase I *C. burnetii* in embryonated eggs or tissue culture results in LPS molecules with decreasing molecular weights, culminating in the severely truncated LPS of avirulent phase II organisms. Phase II LPS contains lipid A and some core sugars but is missing O-antigen sugars and appears to represent the minimal LPS structure of *C. burnetii* (1, 30, 74). Two cloned LPS variants of the virulent Nine Mile phase I (NMI) isolate have been described: Nine Mile Crazy (NMC; intermediate virulence), producing an intermediate-length LPS, and Nine Mile phase II (NMII; avirulent), producing a severely truncated phase II LPS (30, 43). NMII and NMC have large chromosomal deletions (26 and 32 kb, respectively) that eliminate open reading frames (ORFs) involved in the biosynthesis of O-antigen sugars, including the rare sugar virenose (6-deoxy-3-C-methylglucose) (35, 76).

The availability of genome sequences of pathogenic bacteria allows a rapid assessment of intrastrain/species whole genome sequence variation by using comparative genome hybridization (CGH) on DNA microarrays. Comparisons have been made between strains of the same species (18, 50, 56) and between different species of the same genus that differ in virulence potential, environmental origin, and animal host (22, 48). These studies provide insight into pathogen genetic diversity, strain evolution, epidemiology of disease, and virulence potential. Moreover, CGH can be used for high-resolution molecular strain discrimination. Indeed, current serological, PCR, and cell culture-based diagnostic methods fail to discriminate between different *C. burnetii* isolates (21). The genome-wide variations identified by CGH allow a more definitive analysis of the evolutionary relatedness of bacteria than other phylogenetic methods, such as those that rely on variation in the nucleotide sequence of a limited number of conserved genes (e.g., rRNA and housekeeping genes), RFLP profiles, or diversity in variable-number tandem repeats (37). These methods generally do not resolve the loss of DNA and, in the case of RFLP profiles, a single band in one isolate can correspond to multiple bands of different isolates, showing that loss of a restriction site does not indicate the deletion of a whole DNA fragment and its encoded genes (77).

The *C. burnetii* NMI (RSA493) genome has recently been sequenced and consists of a circular chromosome of 1,995,275 bp and a plasmid (QpH1) of 37,393 bp that encode 2,095 and 40 ORFs, respectively (63). Analysis of coding ORFs indicates that, although *C. burnetii* shares similarities in lifestyle and parasitic strategies to other obligate intracellular bacteria, it generally differs considerably with respect to genome size, metabolic and transporter capabilities, the presence of mobile genetic elements, and the extent of genome reduction (23, 55, 57, 63, 84). To further define the genetic diversity of *C. burnetii*, gDNA of 24 isolates of diverse geographical and environmental origins were hybridized to a custom Affymetrix microarray chip (RMLchip_a) containing probe sets corresponding to all ORFs of the NMI isolate. We present here the first *C. burnetii* whole-genome phylogeny. Although the genomes of *C. burnetii* isolates demonstrate considerable relatedness, some isolates are missing a number of ORFs relative to NMI that likely result in different biological properties. Moreover, the results of the present study expand our understanding of the genetics of LPS phase variation and establish a new method for classifying unknown *C. burnetii* isolates.

MATERIALS AND METHODS

***C. burnetii* isolates, cultivation, and purification.** The *C. burnetii* isolates used in the present study are part of the culture collections maintained at the Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, Hamilton, MT, and Texas A&M University System, College Station, TX. Isolate name, geographical/sample origin, plasmid type, clinical disease, and passage history are listed in Table 1. Q321 is a cow's milk isolate from Russia (75), and Dugway 5G61-63 is a tick isolate from Dugway, Utah (68). Le Bruges was isolated in France from an unknown source and BDT 1 was isolated in the United States from a brown dog tick and has not been previously described. The remaining isolates have been previously described (33). *C. burnetii* was propagated in African green monkey kidney (Vero) fibroblasts (CCL-81; American Type Culture Collection) grown in RPMI medium (Invitrogen) supplemented with 2% fetal bovine serum, or in embryonated hen's eggs, as previously described (59). Organisms were purified from infected eggs or Vero cells by Renografin or sucrose density gradient centrifugation (29, 59).

C. burnetii replicative vacuoles were harvested from infected Vero cells on 12-mm glass coverslips by micromanipulation. At 5 days postinfection, coverslips were transferred to a 90-cm diameter petri dish containing approximately 10 ml of RPMI medium. Individual vacuoles were extracted from an infected monolayer by using a Micromanipulator 5171 (Eppendorf), a CellTram vario manual microinjector (Eppendorf), and TransferTips MDS (Eppendorf). The micromanipulator was mounted to the stage of a Zeiss Axiovert 25 inverted microscope, and vacuole harvest was conducted at $\times 200$ magnification. The fluid contents of harvested vacuoles (ca. 1 μ l) were added to 40 μ l of K-36 buffer (0.1 M KCl, 0.015 M NaCl, 0.05 M K_2HPO_4 , 0.05 M KH_2PO_4 [pH 7.0]) in a 1.5-ml microfuge tube, and the mixture was frozen at -80°C .

Genomic DNA isolation and amplification. Total gDNA was purified directly from purified *C. burnetii*, infected guinea pig spleen homogenate (BDT isolate), and infected yolk sac (G isolate) by one of two methods. The first method used an UltraClean microbial DNA isolation kit (MoBio Laboratories, Inc.) according to the protocol recommended by the supplier, with an additional heating step (85°C for 10 min) before physical disruption of the bacterial cells. The second method has been previously described by Samuel et al. (58). All DNA was resuspended in distilled H_2O and frozen at -20°C . Whole-genome amplification (WGA) was conducted on total DNA yielded from 12.5 μ l of infected yolk sac and directly on 1 μ l of *C. burnetii* vacuole suspension without prior DNA purification. Amplifications used a GenomiPhi DNA Amplification Kit (Amersham Biosciences). Specifically, 1 μ l of DNA solution or *C. burnetii* vacuole suspension was added to 9 μ l of sample buffer, which was heated to 95°C for 3 min and then cooled to 4°C on ice. A total of 9 μ l of reaction buffer and 1 μ l of enzyme mix were then added to the cooled DNA sample. The samples were incubated at 30°C for 18 h, followed by inactivation of the Phi29 DNA polymerase by heating at 65°C for 10 min. The samples were cooled on ice and stored at -20°C . Four separate WGA reactions were conducted on DNA from infected yolk sacs, while

TABLE 1. *C. burnetii* isolates used in this study

Genomic group ^a	Plasmid type ^b	Isolate	Isolate designation ^c	Phase ^d	Original source, yr	Disease ^e	Passage history ^f
I	QpH1	Nine Mile RSA493	NMI	I	Montana, tick, 1935	—	307GP/1TC/1EP
I	QpH1	Nine Mile RSA439 (clone 4)	NMII	II	Montana, tick, 1935	—	90EP/1TC/4EP
I	QpH1	Nine Mile RSA514 Crazy	NMC	I/II	Montana, tick, 1935	—	4EP/306GP/3EP
I	QpH1	Australia QD RSA425 (clone 3)	Au	II	Australia, human blood, ~1939	Acute	177EP/1TC/2EP/1VC
I	QpH1	D RSA345	D	II	United states, human blood, 1938	Acute	81EP
I	QpH1	Turkey RSA333	T	II	Turkey, human blood, 1948	Acute	31EP
I	QpH1	African RSA334	Af	I	Central Africa, human blood, 1949	Acute	3HP/4EP
I	QpH1	Giroud RSA431	Gi	I	Central Africa, human blood, 1949	Acute	2GP/2EP
I	QpH1	EI Tayeb RSA342	EIT	I	Egypt, tick, 1967	—	4GP/2EP
I	QpH1	Panama RSA335	Pa	I	Panama, chiggers, 1961	—	4EP
I	QpH1	California 33 RSA329	Ca 33	I	California, cow's milk, 1947	Persistent	6EP
I	QpH1	California 16 RSA350	Ca 16	II	California, cow's milk, 1947	Persistent	38EP
I	QpH1	Ohio 314 RSA270	Oh	I	Ohio, cow's milk, 1956	Persistent	4EP
II	QpH1	Henzerling RSA331	He	II	Italy, human blood, 1945	Acute	36EP
III	QpH1	Idaho Goat Q195	Id	I	Idaho, goat, 1981	Abortion	2EP
IV	QpRS	MSU Goat Q177 (Priscilla)	MSU	I	Montana, goat cotyledon, 1980	Abortion	2EP
IV	QpRS	K Q154	K	I	Oregon, human heart valve, 1976	Endocarditis	2EP
IV	QpRS	P Q173	P	I	California, human heart valve, 1979	Endocarditis	2EP
V	IP	G Q212	G	I	Nova Scotia, human heart valve, 1981	Endocarditis	2EP
V	IP	S Q217	S	I	Montana, human liver biopsy, 1981	Hepatitis	2EP
V	IP	Ko Q229	Ko	I	Nova Scotia, human heart valve, 1982	Endocarditis	2EP
VI	QpDG	Dugway 7E9-12	Du 7E9-12	I	Utah, rodents, 1958	—	3EP
UN	QpDV	Q321	Q321	ND	Russia, cow's milk	Persistent	?/1EP
UN	PTU	Dugway 5G61-63	Du 5G61-63	ND	Utah, tick (<i>Dermacentor parumapertus</i>), 1958	—	?/1EP
UN	PTU	BDT 1	BDT	ND	United States, brown dog tick (<i>Rhipicephalus sanguineus</i>), 1951	—	?/1GP
UN	PTU	Le Bruges	Le B	ND	France, unknown source	—	?/1EP

^a As previously defined by Hendrix et al. (33). UN, unassigned genomic group.

^b As previously described (59) with the exception of QpDV, which was described by Valkova and Kazar (75). IP, integrated plasmid; PTU, plasmid type unknown.

^c Designation of isolates used in this study.

^d Phase defined by complement block titration, M. G. Peacock, Rocky Mountain Laboratories; ND, phase was not determined.

^e Disease associated with each isolate. —, No disease association.

^f The preceding number indicates laboratory passages in guinea pig (GP); tissue culture (TC); eggs (EP); and hamsters (HP). ?, Passage prior to receipt in the author's laboratory unknown.

eight separate reactions were conducted on the *C. burnetii* vacuole suspension. The respective reactions were pooled, and DNA was purified by using *n*-butanol. Briefly, 30 μ l of sterile distilled H₂O and 500 μ l of *n*-butanol was added to each sample, and the mixture was vortexed for 5 s. WGA DNA was pelleted by centrifugation at 12,000 \times g for 10 min, air dried, resuspended in 80 μ l of distilled H₂O, and then quantified spectrophotometrically.

Affymetrix array design. A custom Affymetrix GeneChip (RMLchip_a) was utilized in the present study (39). The RMLchip_a is an antisense oligonucleotide expression array (18- μ m feature size) consisting of ~249,690 25-mer probe pairs, with each *C. burnetii* ORF represented by 16 probe pairs that consist of a perfect-match (PM) probe and a mismatch (MM) probe. The MM probe has a single substitution at position 13 relative to the PM probe. To facilitate the analysis of *C. burnetii* DNA with potential host cell DNA contamination, during the chip design phase all probe set sequences were pruned against human, rat, and mouse genomes. The same pruning process was conducted to prevent cross-hybridization to the 20 additional bacterial genome sequences on the RMLchip_a. The *C. burnetii* component of the RMLchip_a consists of probe sets specific for 2,024 full-length genomic ORFs (1,988 chromosomal ORFs and 36 plasmid ORFs) of the sequenced Nine Mile isolate (RSA493) (63). Probe sets specific for 79 disrupted ORFs that contain deletions or point mutations, resulting in premature stop codons or frameshifts, were also included. Probe sets specific for one representative of the 21 IS1111A, 5 IS30, and 3 IS4s1 IS element families were also included. The total coding region of the *C. burnetii* genome is 1,705,446 bp. The probe sets on the RMLchip_a encompass ca. 30% (511,725 bp) of this region.

Microarray sample preparation and hybridization conditions. Microarray hybridizations were conducted with 7 and 30 μ g of gDNA extracted from purified *C. burnetii* and infected guinea pig spleen homogenate (BDT isolate), respectively. We also used 7 and 10 μ g of pooled WGA DNA samples from *C. burnetii* vacuole suspensions and infected yolk sacs, respectively. Total DNA was fragmented with DNase I (Roche Applied Science; 0.001 U/ μ g of DNA) for 10 min at 37°C in DNase I buffer. DNase I was denatured by heating (10 min at 98°C) and 1 μ l of the fragmented gDNA was electrophoretically analyzed with an Agilent 2100 Bioanalyzer (Agilent Technologies, Inc.) to confirm the generation of DNA fragments between 50 and 200 bp. Fragmented DNA was then 3' end labeled with biotin-ddUTP for 60 min at 37°C by using a BioArray terminal labeling kit (Enzo Life Sciences, Inc.). The fragmented, end-labeled cDNA was

added to hybridization solution that was used to probe the RMLchip_a. Target hybridizations, RMLchip_a washing, staining, and scanning were conducted according to the manufacturer's suggestions at the National Institute of Allergy and Infectious Diseases Affymetrix Core Facility (SAIC-Frederick, Inc.). Prior to hybridization to the RMLchip_a, the gDNA samples in hybridization solution were heated to 95°C for 5 min and then incubated for 5 min on ice, followed by equilibration to room temperature. Hybridization was carried out for 16 h at 50°C with rotation in a hybridization oven (Affymetrix) at 60 rpm. After removal of the hybridization solution, the RMLchip_a was washed and stained using a Fluidics Station 400 according to directions stated by Affymetrix for the GeneChip *Pseudomonas aeruginosa* Array. Labeled RMLchip_a's were then scanned by using an Affymetrix GeneChip Scanner 3000.

Microarray data analysis. Images from scanned microarrays were processed with Affymetrix GeneChip Operating Software (GCOS) version 1.2 (Affymetrix). One to four hybridizations per isolate were performed with separately labeled gDNA or WGA DNA. The difference in intensity value for each probe pair was calculated by subtracting the MM probe intensity value from the PM probe intensity value. The average difference in intensity values for the entire 16 probe pairs for each ORF was calculated by using the statistical expression algorithm incorporated into GCOS. Normalization between arrays was achieved by implementing array-specific scaling factors that were calculated as a ratio of 500 divided by the average of the average difference in intensity values for all ORFs represented on the RMLchip_a. For each ORF, the average difference in intensity was multiplied by the array specific scaling factor. Using Partek Pro software version 6.04.1208 (Partek), sample/reference ratios of signal intensity were calculated and transformed to logarithm base 2. The ratios were then normalized by using the median log₂ ratios of zero, and the final value was calculated from the average of all replicates. ORFs that had a log₂ ratio of less than -1 were considered absent and subjected to PCR and sequencing validation.

Microarray validation. PCR validation was conducted for all ORFs indicated as absent by microarray. The PCR strategy used is shown in Fig. S1 in the supplemental material. PCRs utilized primers specific to ORFs flanking putative deletions. They were first used to amplify the two ORFs immediately flanking polymorphic regions to determine whether potential deletions extended into the flanking ORFs. PCR amplifications were then conducted across polymorphic regions to determine whether ORFs within this region were deleted, contained

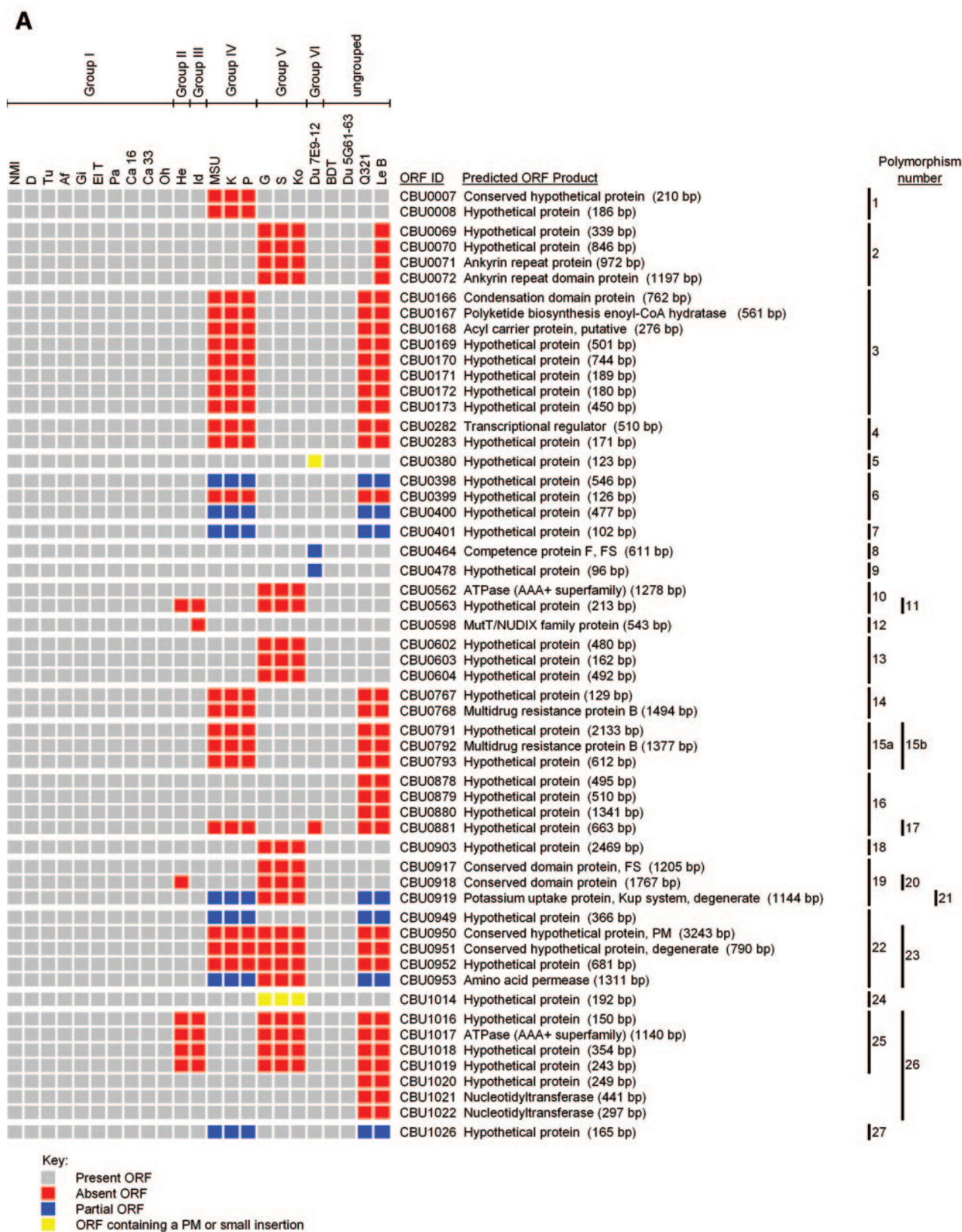


FIG. 1. Comparison of the genomes of 22 *C. burnetii* isolates with the genome of the Nine Mile (RSA493) reference isolate. Labeled DNA was hybridized to the RMLchip_a containing probe sets specific for all ORFs of the Nine Mile isolate. ORF polymorphisms were identified as ORF probe sets showing \log_2 hybridization ratios lower than -1 that were validated as described in Materials and Methods. Red ORFs were completely deleted, blue

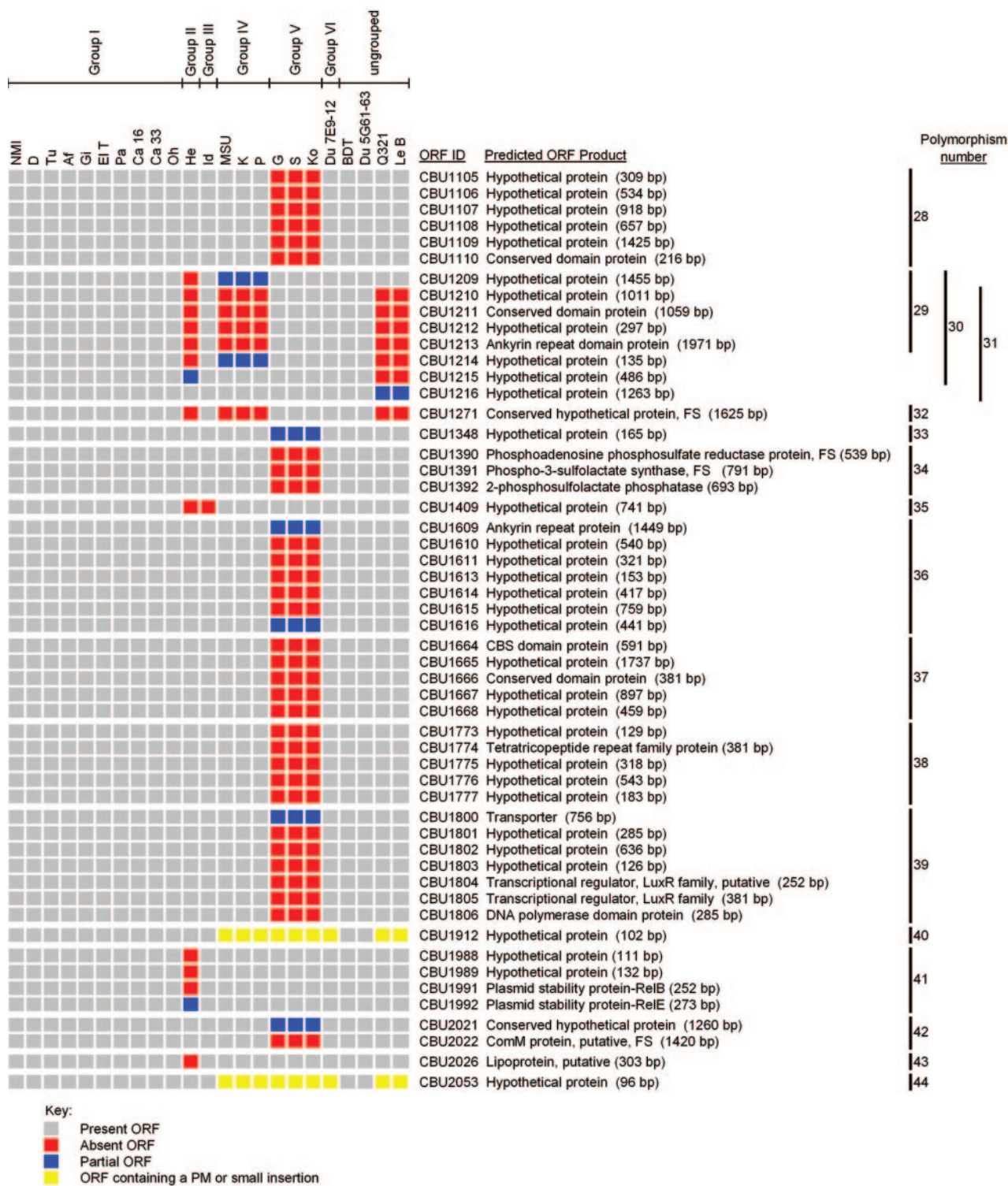


FIG. 1—Continued.

ORFs were partially deleted, and yellow ORFs contained point mutation(s) or small insertions. The number and location of each polymorphism, consisting of single or multiple ORFs, is shown on the right. (A) Chromosomal polymorphisms 1 through 44. (B) Plasmid polymorphisms 45 through 51. The number and location of each polymorphism, consisting of either single or multiple ORFs, is shown on the right. Predicted nonfunctional NMI ORFs containing point mutations (PM) or frameshift (FS) mutations are indicated.

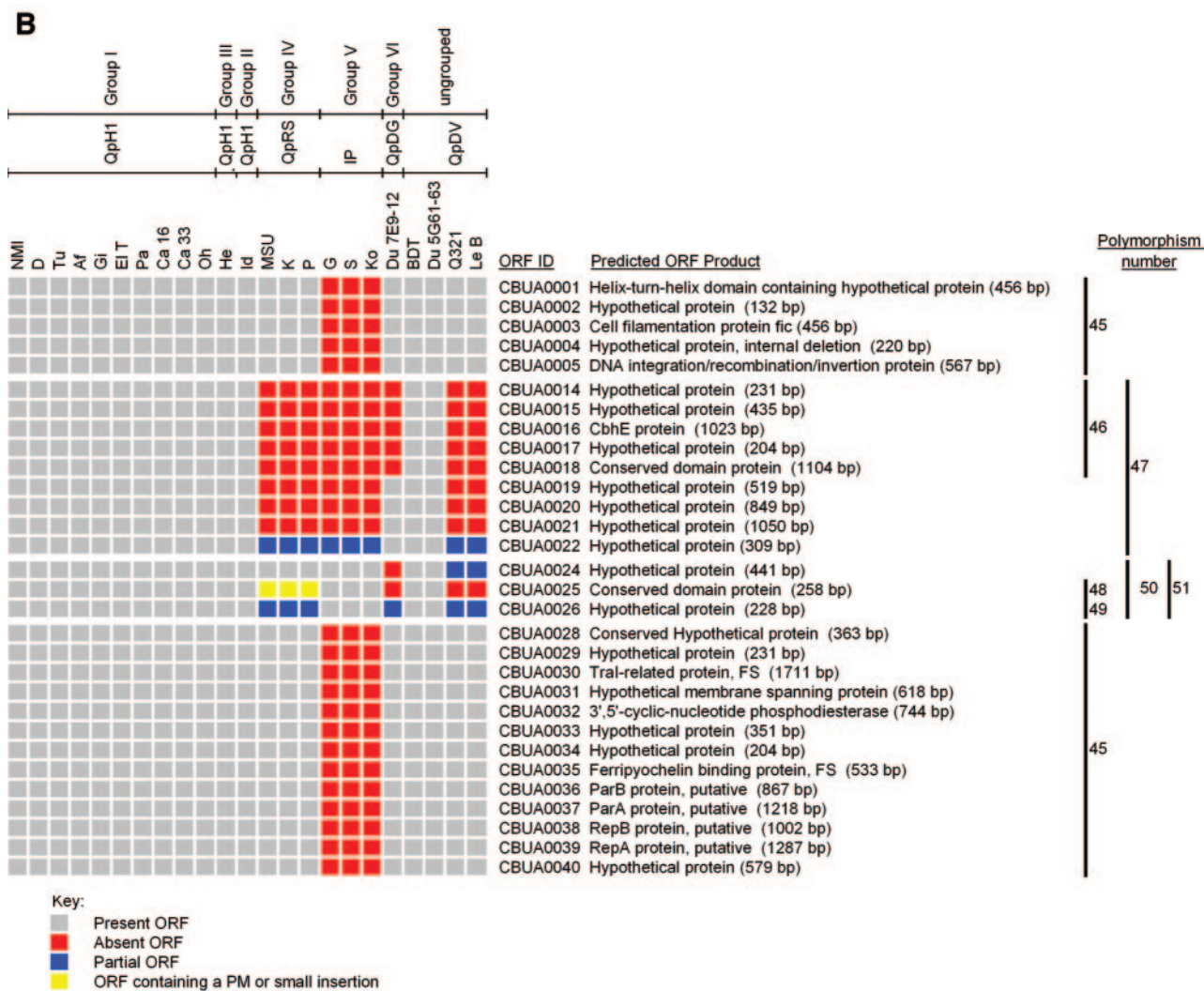


FIG. 1—Continued.

inserted DNA, or were intact. Amplification of NMI genomic DNA was conducted as a control. The ORF-specific primers used in the present study are listed in Table S1 in the supplemental material. Primer sequences were based on the published *C. burnetii* Nine Mile (RSA493) sequence (63) and were manufactured by Sigma Genosys, LP. PCRs (50 µl) containing 100 ng of gDNA were performed by using either AccuPrime *Pfx* DNA polymerase (Invitrogen Corp.) for the predicted short PCR products (<4 kb) or Accuprime *Taq* DNA polymerase High Fidelity (Invitrogen Corp.) for the predicted long PCR products (>4 kb). PCR products analyzed by DNA sequencing were processed by using a QIAquick PCR purification kit (QIAGEN) and sequenced by the RML Genomics Core Facility (Rocky Mountain Laboratories, Hamilton, MT). DNA sequencing reactions were performed with an ABI Prism Dye Terminator Cycle Sequencing Ready Reaction sequencing kit according to the manufacturer's instructions (Applied Biosystems, Inc.), and reactions were analyzed by using a model 3700 automated DNA sequencer (Applied Biosystems, Inc.).

Phylogenetic analysis. A numerical score was assigned to each isolate ORF based on the following criteria: 1 for present, 2 for partially deleted, 3 for absent, and 4 for containing a point mutation or small insertion (<50 bp). ORF designations were then analyzed by using PAUP software version 4.0b10 (PPC; Sinauer Associates, Inc.). Phylogenetic trees were constructed by using maximum-parsimony with equal weighting following a heuristic search.

LPS extraction and analysis. *C. burnetii* LPS was extracted by a modified hot phenol method that exploits the molecule's hydrophilic nature (30). Briefly, 1 mg (dry weight) of purified *C. burnetii* was suspended in 1 ml of 50% phenol. The sample was boiled for 10 min, incubated on ice for 5 min, and then centrifuged

at 14,000 × *g* for 5 min. The aqueous phase was collected, and the extraction was repeated on the pellet. The aqueous phases from both extractions were pooled (ca. 1 ml) and vacuum dried overnight. The resulting pellet was dissolved in 100 µl of distilled H₂O, the sample was dried again, and the pellet was dissolved in 50 µl of sample buffer. Samples were then separated on a 12.5% sodium dodecyl sulfate-polyacrylamide gel. LPS was stained by using a SilverQuest Silver staining kit (Invitrogen Corp.).

Deposition of microarray data in the public databases. The microarray data have been deposited in the microarray database at European Bioinformatics Institute (<http://www.ebi.ac.uk>) under the accession number E-TABM-35.

RESULTS

CGH microarray analysis. We used microarray-based CGH as a highly resolving method to explore the genetic diversity of *C. burnetii* isolates from disparate sources. Beginning with sequence information compiled by The Institute for Genomic Research during the genomic sequencing of *C. burnetii* NMI, a custom Affymetrix GeneChip (RMLchip_a) was constructed that contains probe sets specific for 2,103 predicted ORFs (2,063 chromosomal and 40 plasmid ORFs) of *C. burnetii* (63). Seventy-nine of these ORFs are considered nonfunctional due to

TABLE 2. ORF polymorphisms of 22 *C. burnetii* isolates relative to the Nine Mile reference isolate

Isolate	Chromosomal ORF polymorphism type ^a				Plasmid ORF polymorphism type ^b				% of ORFs with polymorphisms ^g
	No. of single-event polymorphisms ^c	Point/insertion polymorphism ^d	Partial ORFs ^e	Deleted ORFs ^f	No. of single-event polymorphisms ^c	Point/insertion polymorphism ^d	Partial ORFs ^e	Deleted ORFs ^f	
NMI									0
D									0
Tu									0
Af									0
Gi									0
EI T									0
Pa									0
Ca 16									0
Ca 33									0
Oh									0
Id	4			7					0.3
He	8		2	18					0.9
MSU	15	2	9	27	3	2	1	8	2.3
K	15	2	9	27	3	2	1	8	2.3
P	15	2	9	27	3	2	1	8	2.3
G	18	3	4	53	2		1	26	4.1
S	18	3	4	53	2		1	26	4.1
Ko	18	3	4	53	2		1	26	4.1
Du 7E9-12	6	3	2	1	2		1	7	0.7
BDT									0
Du 5G61-63									0
Q321	15	2	8	37	2		3	9	2.8
Le B	16	2	8	41	2		3	9	3.0

^a Type of polymorphism associated with any of the 2,063 NMI chromosomal ORFs.
^b Type of polymorphism associated with any of the 40 NMI QpH1 plasmid ORFs.
^c Predicted number of separate genetic events resulting in ORF polymorphism within each isolate.
^d Number of ORFs of <250 bp presumably containing point mutations or small insertions, where the PCR indicates the entire ORF sequence is present.
^e Number of partially deleted ORFs.
^f Number of completely deleted ORFs.
^g Percentage of NMI ORFs that are polymorphic.

the presence of internal frameshifts or premature stop codons (63). NMI is considered a reference isolate of *C. burnetii* that is biologically representative of acute disease isolates (33). It is the original U.S. isolate that was obtained from a tick in Montana in 1935 (17). The genomes of 23 *C. burnetii* isolates and two passage history variants of NMI, having different biological properties and obtained from geographically diverse areas (Table 1), were compared to NMI by hybridization to the RMLchip_a. Isolates include those obtained from acute and chronic human disease (endocarditis or hepatitis), goat placenta following abortion, cow's milk, rodents, ticks, and chiggers. Nineteen of these isolates reside in genomic groups I to VI according to the nomenclature of Hendrix et al. (33). Four ungrouped isolates, i.e., Q321, Le Bruges, Dugway 5G61-63, and BDT 1, were also included in the present study.

Figure 1 depicts chromosomal (Fig. 1A) and plasmid (Fig. 1B) ORF polymorphisms relative to NMI that were present in individual *C. burnetii* isolates as identified by CGH. In the present study a polymorphism is defined as a single or multiple linked ORF(s) where negative hybridization occurred due to complete or partial ORF deletion, point mutation(s), or DNA insertion. The number of polymorphic ORFs ranged from 0 in genomic group I isolates to 87 in genomic group V isolates, which represents 4.1% of the NMI coding capacity (Table 2). Between all isolates, a total of 139 ORF polymorphisms (109 chromosomal and 30 plasmid) were identified. A total of 34 of these ORFs have putative functions (Table 3), while 93 encode hypothetical proteins. The remaining 12 polymorphic ORFs

are annotated but are predicted to be nonfunctional because of a frameshift or truncation. The number of predicted single-event polymorphisms, i.e., changes predicted to occur once but affecting either single or multiple ORFs, varied between isolates and ranged from 0 in genomic group I isolates to 20 in genomic group V isolates. The size of polymorphisms ranged from a single ORF (e.g., CBU0478 in Du 7E9-12) to 18 ORFs (e.g., plasmid ORFs CBUA0028 to CBUA0005 in genomic group V isolates with integrated "plasmid-like" sequences). Isolates within previously described genomic groups all had the same polymorphisms, confirming their relatedness based on previous RFLP studies (33).

Polymorphisms within genomic groups I to IV included six chromosomal polymorphisms (polymorphisms 17, 25, 32, 35, 40, and 44) common to at least two genomic groups. Four chromosomal regions of difference (RD), defined as regions that contain polymorphisms of different sizes within the same relative chromosomal locations, were also observed (polymorphisms 10 and 11, 19 to 21, 22 and 23, and 29 and 30). The remaining 26 chromosomal polymorphisms were unique to a single genomic group. No polymorphisms were observed with the previously ungrouped isolates BDT and Du 5G61-63, indicating that these isolates reside within genomic group I. In contrast, ungrouped isolates Q321 and Le B contained a large number of predicted single-event chromosomal polymorphisms (15 and 16, respectively). They were identical apart from a single polymorphism encompassing CBU0069 to CBU0072 that is present in Le B but not Q321. This polymor-

TABLE 3. Nine Mile ORFs with a predicted function that are completely or partially deleted in at least one *C. burnetii* isolate

ORF type and ID ^a	Description	Size (bp)
Chromosomal ORFs		
CBU0071	Ankyrin repeat protein	972
CBU0072	Ankyrin repeat domain protein	1,197
CBU0166	Condensation domain protein	762
CBU0167	Polyketide biosynthesis enoyl-CoA hydratase ^b	561
CBU0168	Acyl carrier protein, putative	276
CBU0282	Transcriptional regulator	510
CBU0562	ATPase (AAA ⁺ superfamily)	1,278
CBU0598	MutT/NUDIX family protein	543
CBU0768	Multidrug resistance protein B	1,494
CBU0792	Multidrug resistance protein B	1,377
CBU0953	Amino acid permease	1,311
CBU1017	ATPase (AAA ⁺ superfamily)	1,140
CBU1021	Nucleotidyltransferase	441
CBU1022	Nucleotidyltransferase	297
CBU1213	Ankyrin repeat domain protein	1,971
CBU1392	2-phosphosulfolactate phosphatase	693
CBU1609	Ankyrin repeat protein	1,449
CBU1664	CBS domain protein	591
CBU1774	Tetratricopeptide repeat family protein	381
CBU1800	Transporter	756
CBU1804	Transcriptional regulator, LuxR family, putative	252
CBU1805	Transcriptional regulator, LuxR family	381
CBU1806	DNA polymerase domain protein	285
CBU1991	Plasmid stability protein-RelB	252
CBU1992	Plasmid stability protein-RelE	273
CBU2026	Lipoprotein, putative	303
Plasmid ORFs		
CBUA0003	Cell filamentation protein fic	456
CBUA0005	DNA integration/recombination/inversion protein	567
CBUA0016	CbhE protein	1,023
CBUA0032	3',5'-cyclic-nucleotide phosphodiesterase	744
CBUA0036	ParB protein, putative	867
CBUA0037	ParA protein, putative	1,218
CBUA0038	RepB protein, putative	1,002
CBUA0039	RepA protein, putative	1,287

^a ID, identification number.

^b CoA, coenzyme A.

phism was also observed in genomic group V isolates. Of the 15 chromosomal polymorphisms conserved between Le B and Q321, 12 were also present in genomic group IV isolates. Five of these polymorphisms (21, 22, 32, 40, and 44) were also detected in isolates of additional genomic groups, with polymorphisms 21 and 22 included in chromosomal RDs. The remaining polymorphisms (16, 26, and 31) were uniquely sized and resulted in three additional chromosomal RDs.

Seven polymorphisms (45 to 51) relative to the QpH1 plasmid of NMI were observed. One polymorphism (45) was unique to genomic group V isolates that have plasmid-like sequences integrated into the chromosome. A plasmid RD was observed that includes polymorphism 47 of genomic groups IV, V, Le B, and Q321 and polymorphism 46 of Du 7E9-12, which is a subregion of polymorphism 47. A second plasmid RD is represented by the remaining plasmid polymorphisms (48 to 51) that were observed in genomic groups IV, VI, and the ungrouped isolates Le B and Q321. Although the plasmid content of Le B is unknown, it likely carries QpDV since the plasmid polymorphisms of this isolate were identical to Q321,

TABLE 4. Conserved ORFs of QpH1, QpRS, QpDV, QpDG, and integrated plasmid sequences

Gene ID ^a	Predicted protein function
CBUA0006	Hypothetical protein
CBUA0007	Hypothetical protein
CBUA0008	Phage tail assembly protein
CBUA0009	Nemacidal toxin, putative, degenerate
CBUA0010	Site-specific recombinase, phage integrase family
CBUA0011	Probable integrase/recombinase ripX
CBUA0012	Hypothetical protein
CBUA0013	Hypothetical protein
CBUA0023	Hypothetical protein
CBUA0027	DNA-binding protein

^a ID, identification number.

which contains this plasmid (75). Only 10 ORFs were conserved between all plasmid types and integrated plasmid-like sequences (Table 4), with one (CBUA0009) predicted to be nonfunctional in NMI.

Microarray validation. To experimentally validate polymorphisms identified by CGH, PCR (with or without DNA sequencing of PCR replicons) of polymorphic regions was conducted on all ORFs indicated as absent. Figure S1 in the supplemental material depicts the PCR strategy used to characterize polymorphisms, and Table 5 summarizes the validation results. In general, deletions varied widely in size, ranging from 43 bp (CBU0478 in Du 7E9-12) to 19.9 kb (CBUA0028 to CBUA0005 in the genomic group V isolates). The majority of polymorphisms consisted of completely deleted individual or linked ORFs with the deletion not extending into flanking ORFs. Polymorphisms that consisted of partially deleted ORFs, either at the boundary of a multiple ORF polymorphism or as a single ORF polymorphism, were also identified (e.g., polymorphisms 6 and 8, respectively). Some PCR products were larger than predicted for a deleted ORF(s), indicating the presence of inserted DNA, which in some cases (e.g., polymorphisms 1, 4, and 22) was confirmed by DNA sequencing. PCR products indicating the presence of full-length ORFs that were called absent by microarray were obtained for 5 of 10 ORFs of <250 bp (polymorphisms 5, 24, 40, 44, and 48). DNA sequencing of PCR products specific to polymorphisms 5, 40, and 48 revealed point mutations that presumably prevented hybridization of gDNA to the predominantly overlapping probe pairs corresponding to these smaller ORFs. Polymorphism 5 (CBU0380) contained a synonymous change, whereas polymorphisms 40 (CBU1912) and 48 (CBUA0025) contained nonsynonymous changes resulting in frameshifts that are predicted to disrupt gene function. Although not confirmed by DNA sequencing, point mutation(s) in CBU1014 and CBU2053 also likely explain polymorphisms 22 and 44, respectively. PCR validation identified two cases (polymorphisms 4 and 30) where flanking genes (CBU0281, CBU0284, and CBU1208) contained small partial deletions not detected by CGH. In all cases examined, sequencing of PCR products revealed that the deletion boundaries of a given polymorphism were the same for isolates within the same genomic group. Between genomic groups, deletion boundaries could vary (e.g., polymorphisms 19 to 21 in genomic groups II, IV, and V). Validation results also indicated a high degree of genome synteny between *C. burnetii* isolates as PCR products were obtained for all polymorphisms.

TABLE 5. Characterization of genomic polymorphisms

Polymorphism no.	Polymorphic region ^a	Size of corresponding region in NMI (bp) ^b	PCR/sequencing results ^c
1	CBU0007 to CBU0008	1,785	Insertion of ~7 kb of DNA adjacent to transposase (CBU0006); PCR products were not obtained for CBU0007 or CBU0008
2	CBU0069 to CBU0072	3,909	~2.7-kb deletion with flanking IS1111A transposon-like sequence ^{d,f}
3	CBU0166 to CBU0173	4,739	~3.2-kb deletion and insertion of unknown DNA ^{d,f}
4	CBU0282 to CBU0283	1,482	Large insertion (>4 kb) that eliminates CBU0282, CBU0283, the first four codons of CBU0284, and the last six codons of CBU0281 and is flanked by 5-bp repeats ^{d,e,f}
5	CBU0380	253	Single synonymous point mutation (TCG to TCC) ^d
6	CBU0398 to CBU0400	1,143	793-bp deletion that eliminates the last 153 codons of CBU0398 and first 102 codons of CBU0400 ^d
7	CBU0401	291	Small deletion of <50 bp
8	CBU0464	627	470-bp deletion that eliminates the first 158 codons ^{d,g}
9	CBU0478	410	43-bp deletion flanked by 6-bp repeat ^{d,e}
10	CBU0562 to CBU0563	3,003	~2.5-kb deletion
11	CBU0563	880	~800-bp deletion
12	CBU0598	606	~600-bp deletion
13	CBU0602 to CBU0604	2,299	2,013-kb deletion that eliminates CBU0602 to CBU0604 ^d
14	CBU0767 to CBU0768	2,476	Large insertion (>7 kb) ^f
15a	CBU0791 to CBU0793	4,371	~2.7-kb deletion in MSU, K and P isolates
15b	CBU0791 to CBU0793	4,371	~2.1-kb deletion in Le B and Q321 isolates
16	CBU0878 to CBU0881	4,878	~3.1-kb deletion adjacent to disrupted transposase (CBU0882)
17	CBU0881	2,321	~1-kb deletion
18	CBU0903	2,524	~2.4-kb deletion
19	CBU0917 to CBU0919	4,869	~3.8-kb deletion
20	CBU0918	4,869	~2-kb deletion
21	CBU0919	1,574	~700-bp deletion
22	CBU0949 to CBU0953	6,954	~5.7-kb deletion and insertion of unknown DNA that eliminates the first 119 codons of CBU0949 ^{d,f}
23	CBU0950 to CBU0953	6,443	~5.9-kb deletion
24	CBU1014	1,029	No deletion detectable, probable point mutation
25	CBU1016 to CBU1019	2,759	2.463-kb deletion that eliminates CBU1016 to CBU1019 and is flanked by 175-bp repeats that are 89% identical ^{d,e}
26	CBU1016 to CBU1022	4,695	4.114-kb deletion that eliminates CBU1016 to CBU1022 and is flanked by 175-bp repeats that are 89% identical ^{d,e}
27	CBU1026	505	Small deletion of less than 50 bp
28	CBU1105 to CBU1110	5,875	~4.5-kb deletion that is adjacent to a transposase (CBU1104)
29	CBU1209 to CBU1213	6,614	5.987-kb deletion that eliminates the first 16 codons of CBU1214, CBU1213 to CBU1210, and the first 295 codons of CBU1209; deleted region is near a transposase (CBU1218) ^d
30	CBU1209 to CBU1215	7,301	7.342-kb deletion that eliminates the last 39 codons of CBU1208, CBU1209 to CBU1214, and the last 137 codons of CBU1215; deleted region is near a transposase (CBU1218) ^d
31	CBU1210 to CBU1216	7,636	6.674-kb deletion that eliminates CBU1210 to CBU1215 and the first 395 codons of CBU1216; deleted region is near a transposase (CBU1218) ^d
32	CBU1271	3,197	1.659-kb deletion with 9-bp flanking repeats that eliminates CBU1271; deleted region is adjacent to a transposase (CBU1270) ^{d,e}
33	CBU1348	508	~500-bp deletion
34	CBU1390 to CBU1392	8,304	~3-kb deletion
35	CBU1409	1,448	1,091-bp deletion flanked by two 104-bp repeats that are 94% identical ^{d,e}
36	CBU1609 to CBU1616	4,720	4.426-kb deletion that eliminates the first 451 codons of CBU1609, CBU1610 to CBU1615, and the last 98 codons of CBU1616 and is flanked by 5-bp repeats ^{d,e}
37	CBU1664 to CBU1668	4,419	~4.2-kb deletion
38	CBU1773 to CBU1777	2,012	~1.7-kb deletion
39	CBU1800 to CBU1806	3,783	3.316-kb deletion that eliminates the first 219 codons of CBU1800 and CBU1801 to CBU1806 and is flanked by 3-bp repeats ^{d,e}
40	CBU1912	337	Two point mutations and insertion of a single base pair ^d
41	CBU1988 to CBU1991	2,430	2.257-kb deletion that eliminates CBU1988 to CBU1991 (containing the transposase CBU1990) and the first 27 codons of CBU1992 and is flanked by two 49-bp repeats that are 92% identical ^{d,e}
42	CBU2021 to CBU2022	2,835	1,881-bp deletion that eliminates the majority of previously disrupted CBU2022 and the first 160 codons of CBU2021, flanked by 8-bp repeats; 83-bp insertion is also present at the start of CBU2022 ^{d,e}
43	CBU2026	745	~400-bp deletion
44	CBU2053	384	No deletion detected, probable point mutation

Continued on following page

TABLE 5—Continued

Polymorphism no.	Polymorphic region ^a	Size of corresponding region in NMI (bp) ^b	PCR/sequencing results ^c
45	CBUA0028 to CBUA0005	13,958	Integration of plasmid genes (CBUA0006 to CBUA0013 and CBUA0023 to CBUA0027) between the chromosomal genes CBU1710 and CBU1711 results in loss of CBUA0028 to CBUA0005 ^g
46	CBUA0014 to CBUA0018	3,947	~3.5-kb deletion
47	CBUA0014 to CBUA0022	7,425	7.0-kb deletion and corresponding 3-bp insertion that eliminates CBUA0014 to CBUA0021 and the first 24 codons of CBUA0022 ^{d,f,g}
48	CBUA0025	748	A single 2-bp insertion and two 1-bp deletions ^d
49	CBUA0026	1,229	3.079-kb insertion containing <i>cbbE</i> at nucleotide 147 ^{d,f,g}
50	CBUA0024 to CBUA0026	2,113	1.278-kb deletion relative to QpH1; QpDV has the same 3.079-kb insertion within CBUA0026 as QpRS; a subsequent 1.802-kb deletion eliminated the last 777 bp of the QpRS deletion, the last 44 codons of CBUA0024, CBUA0025, and the first 48 codons of CBUA0026 ^{d,f}
51	CBUA0024 to CBUA0026	2,113	~700-bp deletion relative to QpH1; QpDV has the same 3.079-kb insertion within CBUA0026 as QpRS; a subsequent ~600-bp insertion eliminated the last 2,676 bp of the 3.079-kb insertion, CBUA0024 and CBUA0025, and the first 48 codons of CBUA0026 ^{d,f,g}

^a Region containing polymorphic DNA as indicated by microarray analysis.
^b Size of the polymorphic region, including flanking intergenic regions, based on the published NMI sequence (63).
^c Polymorphisms were examined by PCR with or without DNA sequencing of the PCR product. Polymorphism sizes were calculated by subtracting the size of the PCR fragment obtained for each polymorphic region from the predicted size of products based on the NMI genome sequence, which was confirmed by PCR of NMI genomic DNA.
^d Polymorphic regions further analyzed by end sequencing of PCR products.
^e Polymorphisms that contain flanking repeat sequences.
^f Polymorphisms that result from insertion of new DNA.
^g Confirmed with published sequence data.

Analysis of a chromosomal and plasmid RD. Sequence analysis of polymorphisms 29 to 31, representing a *C. burnetii* chromosomal RD, suggested this RD was generated by unique deletion events within this region. Polymorphism 29 consisted of an ~6-kb deletion (CBU1209 to CBU1214), while the deletion compromising polymorphism 31 was ~6.7 kb in size (CBU1210 to CBU1216). Polymorphism 30 (CBU1209 to CBU1215), consisted of an ~7.3-kb deletion that had a different 5' deletion point from polymorphism 29.

Polymorphisms 48 to 51 (CBUA0024 to CBUA0026) constitute a plasmid RD shared between QpRS, QpDV, and QpDG. Microarray validation indicated that CBUA0026 contains an insertion of variable size. Specifically, in QpRS CBUA0026 is disrupted by a 3,079-bp insertion. Remnants of the same insertion were present in CBUA0026 of QpDV and QpDG, consisting of 777 and 2,676 bp of the 3' sequence, respectively. Further polymorphisms in this RD were observed in QpDV and QpDG that consisted of deletion of part of CBUA0026, all of CBUA0025, and part (QpDV) or all (QpDG) of CBUA0024. Collectively, these data suggest these and other RDs are hotspots of genomic sequence variation.

Putative functions of polymorphic ORFs. *C. burnetii* encodes a unusually high proportion of hypothetical proteins (63). Accordingly, our analysis showed that 93 (67%) of polymorphic ORFs are annotated as encoding hypothetical proteins. An additional 12 (9%) are annotated as nonfunctional. The remaining polymorphic ORFs have a range of predicted cellular functions, including metabolic control, DNA/RNA processing, nutrient transport, stress response, plasmid stability, and putative host function modulation (Table 3).

Five polymorphic ORFs are predicted to encode proteins containing ankyrin or tetratrico peptide repeats, domains typ-

ically found in eukaryotic proteins that facilitate protein-protein interactions (25, 44). In eukaryotes, proteins with these domains serve a diverse set of functions involving cytoskeletal interactions, transcriptional regulation, signal transduction, and cell cycle regulation (25, 44). Of the 14 ankyrin repeat proteins (ANKs) encoded by the *C. burnetii* genome, 4 (CBU0071, CBU0072, CBU1213, and CBU1609) were polymorphic: 3 in genomic group V and Le B and 1 in genomic groups II, IV, and Q321. In genomic group V isolates, one polymorphic ORF (CBU1774) was predicted to encode a tetratrico peptide repeat protein.

Sixteen polymorphic ORFs were common to genomic groups IV and V and the Q321 and LeB isolates. One of these, CBUA0016 (*cbhE'*), was previously thought to be specific to isolates causing acute disease (42), an idea later discounted upon analysis of a larger group of isolates (65). A second polymorphic ORF (CBU0952) encodes a 28-kDa immunodominant outer membrane protein termed AdaA that was previously found to be synthesized only by acute disease isolates in genomic groups I and II (82). A third polymorphic ORF (CBU0953) encodes an amino acid permease. ORFs encoding 14 amino acid and 3 peptide transporters are present in the *C. burnetii* genome, suggesting functional redundancy within this family of proteins (63).

In addition to ANKs and amino acid/peptide transporters, polymorphisms were identified in seven additional ORFs that are part of large gene families. First, two of at least five predicted nucleotidyltransferases, a family of proteins with roles in DNA repair, transcription, DNA replication and RNA processing, were polymorphic in Q321 and Le B, and one was polymorphic in genomic group IV isolates. Second, 2 of 21 predicted ATPases, proteins associated with a variety of cellu-

lar functions, were polymorphic in genomic group IV, while one was polymorphic in genomic groups II, III, Q321, and Le B. Lastly, two of five ORFs predicted to encode multidrug resistance B proteins were polymorphic in genomic group IV, Q321, and Le B.

The acute disease isolate He contains a unique polymorphism (polymorphism 41) that encompasses *relB* (CBU1991) and the beginning of *relE* (CBU1992). In *Escherichia coli*, these genes regulate the stringent response to amino acid starvation that results in inhibition of translation (11, 12). A second unique He polymorphism (polymorphism 11; CBU0598) is predicted to encode a protein that is part of the MutT/NUDIX (nucleoside diphosphate linked to some other moiety, X) family of proteins, a family consisting of widely distributed enzymes that function by removing potentially hazardous metabolites and/or modulating the buildup of biochemical pathway intermediates (6). Genomic group V isolates contained two polymorphic ORFs with homology to *luxR*. LuxR is a transcriptional activator that acts as a receptor for molecules involved in bacterial quorum sensing (26).

Possible mechanisms of polymorphism generation. Microarray validation indicated that most polymorphisms were generated via DNA insertion or DNA excision and that these events were likely facilitated by flanking repetitive sequences or transposon-like sequences. We divided polymorphisms into five categories based on proximity to repeat sequences or transposases and, if present, the character of inserted DNA. The first category was polymorphisms consisting of deletions that are immediately flanked by repeat sequences. These sequences were typically small (<9 bp), but in four cases large repeats ranging in size from 49 to 175 bp were observed (polymorphisms 25, 26, 35, and 41). Interestingly, the largest of these repeats was a family of three, having ca. 89% identity, with one repeat each located between CBU1015 and CBU1016, CBU1019 and CBU1020, and CBU1022 and CBU1023. In genomic groups II, III, and V, homologous recombination probably occurred between the first two repeats resulting in the deletion of CBU1016 to CBU1019 (polymorphism 25), while recombination between the first and third repeat likely resulted in deletion of CBU1016 and CBU1022 (polymorphism 26) in Q321 and Le B. A second category was polymorphisms consisting of deletions that are less than five ORFs away from a transposase (polymorphisms 1, 16, 28, 29 to 31, and 32). A third category was polymorphisms consisting of deletions where the corresponding DNA in NMI contains a transposase (polymorphism 41), suggesting that the excision of this element in other isolates resulted in this polymorphism. A fourth category was polymorphisms that resulted from DNA insertion. In all but one case, the inserted DNA had no matches to sequences in the National Center for Biotechnology Information database (<http://www.ncbi.nlm.nih.gov/BLAST/>) and presumably represents novel *C. burnetii* sequences. The exception was an insertion of an IS1111A transposase-like sequence in polymorphism 2. The fifth category consisted of “clean” deletions whereby an entire region was missing and with no repeat sequences in the immediate vicinity (e.g., polymorphism 31).

LPS phase variation. The only known difference between virulent phase I and avirulent phase II *C. burnetii* is production of a severely truncated LPS by phase II organisms (2).

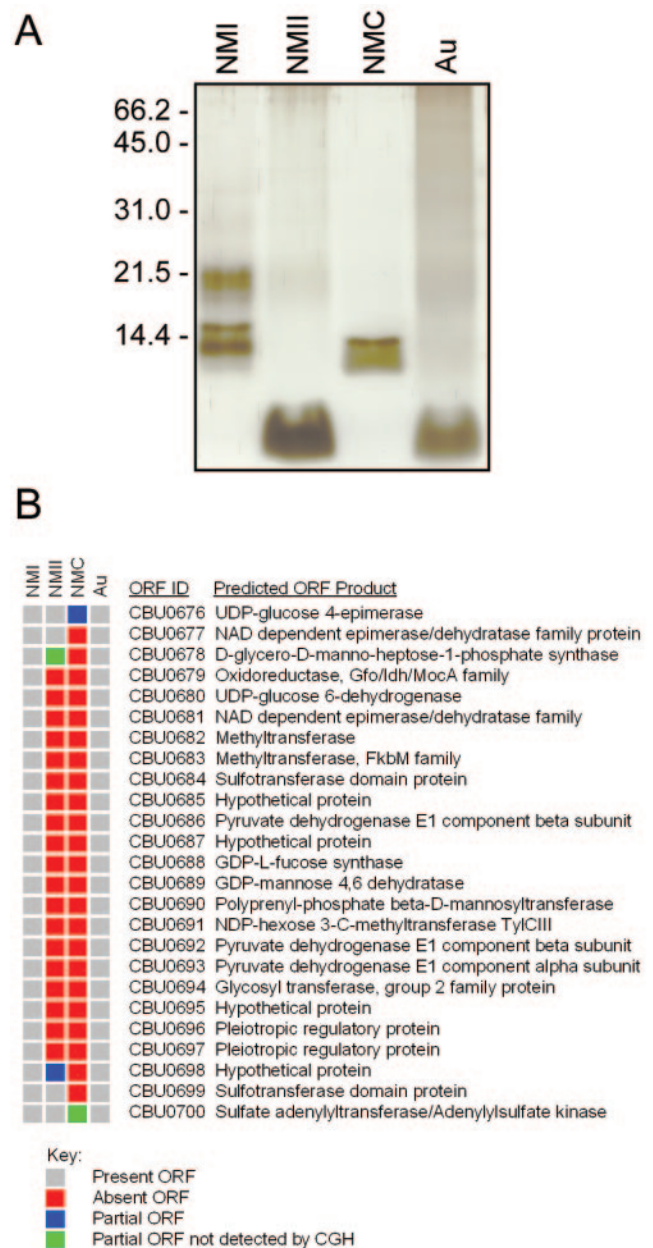


FIG. 2. LPS profiles and ORF polymorphisms of *C. burnetii* isolates synthesizing different LPS chemotypes. (A) Silver-stained sodium dodecyl sulfate–15% polyacrylamide gel electrophoresis profiles of purified LPS molecules from *C. burnetii* Nine Mile I (NMI), Nine Mile II (NMII), Nine Mile Crazy (NMC), and Australia QD (Au) isolates. The relative sizes of molecular mass markers in kilodaltons are shown on the left. (B) Comparison of the genomes of NMII, NMC, and Au isolates with the genome of the Nine Mile (RSA493) reference isolate. Labeled DNA was hybridized to the RMLchip_a containing probe sets specific for all ORFs of the Nine Mile isolate. ORF polymorphisms were identified as ORF probe sets showing log₂ hybridization ratios lower than –1. Red ORFs were completely deleted and blue ORFs were partially deleted. Partial ORFs that were previously identified by Hoover et al. (35), but not detected by the CGH, are depicted in green.

To better understand the genetic basis of LPS phase variation and the attenuated virulence of *C. burnetii* phase variants, we conducted CGH of the genomes of NMC, NMII, and the Australian QD (Au) isolate. The LPS phase varia-

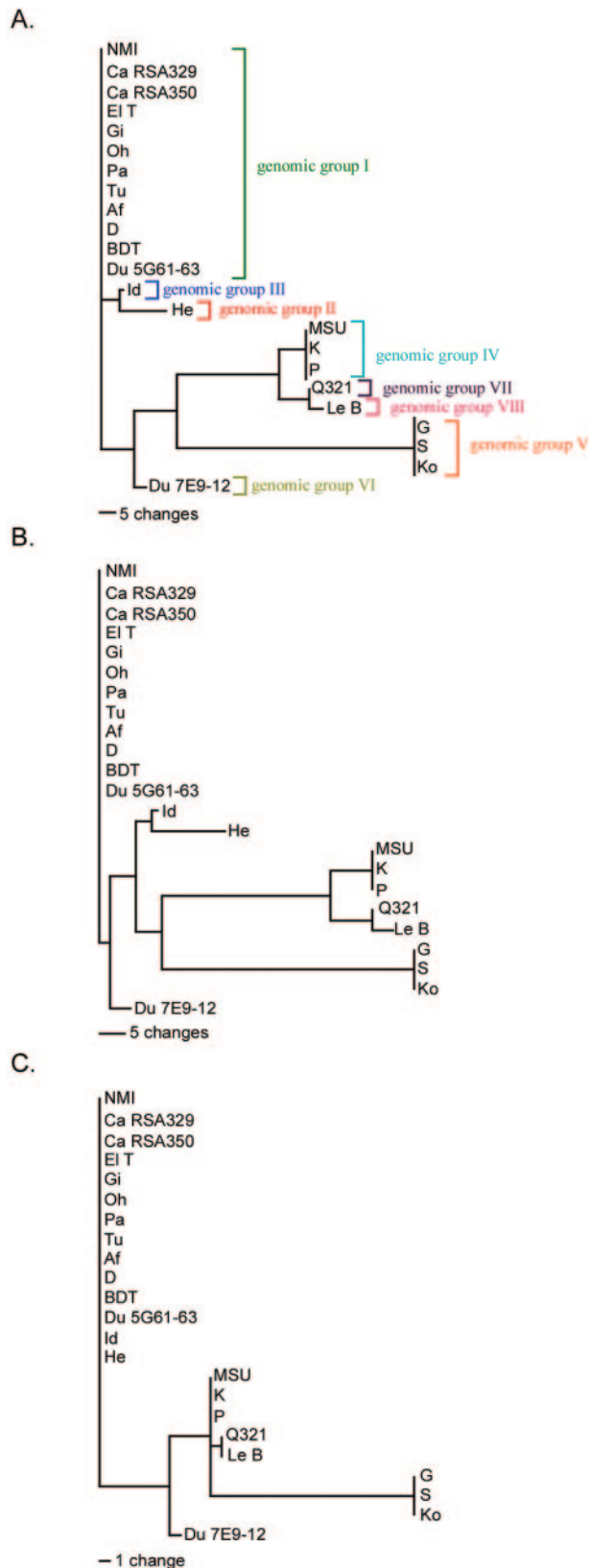


FIG. 3. Parsimony analysis of chromosomal and plasmid ORF content of *C. burnetii* isolates. Phylogenetic trees were constructed by using both chromosomal and plasmid ORFs (with 123 of 132 variable characters being parsimony informative and equally weighted) (A), chro-

tion phenomenon is illustrated in Fig. 2A. NMII produced a severely truncated LPS (~2.5 kDa) relative to the ladderized molecule of NMI (~19, ~15, and ~14.3 kDa). NMC produced an intermediate length LPS (~14.3 kDa). The Au isolate, subjected to 177 egg passages and serologically defined as phase II, produced an LPS with a molecular mass similar to that of NMII (~2.5 kDa). The only NMII and NMC polymorphisms revealed by CGH were 20 ORFs (CBU0679 to CBU0698) and 24 ORFs (CBU0676 to CBU0699), respectively, that are included in the deletions previously described by Hoover et al. (35) (Fig. 2B). No other polymorphisms were detected in either isolate. Interestingly, the Au isolate contained no polymorphisms, indicating that a large deletion is not required for LPS phase variation.

CGH using whole-genome amplified DNA. Obtaining purified *C. burnetii* is an arduous process that requires purification of the organism from cell culture or embryonated eggs in a biosafety level 3 laboratory setting. We therefore tested whether *C. burnetii* DNA derived from a small amount of crude sample was suitable for CGH analysis. Whole-genome amplification was conducted on total DNA extracted from 50 μ l of yolk sac infected with the G isolate and directly on ca. 20% of the content of a *C. burnetii* replicative vacuole containing NMII that was harvested from a Vero cell monolayer by micro-manipulation. Amplified DNA was labeled and hybridized to the RMLchip_a. Hybridization profiles of these samples were identical to those obtained using gDNA extracted from purified NMII and G bacteria (data not shown).

Phylogenetic analysis of *C. burnetii* isolates. The phylogenetic relationships between the isolates used in the present study (excluding NMII and NMC) were analyzed by using PAUP software. Phylogenetic trees were constructed using microarray data based on hybridization signals of all 2,103 chromosomal and plasmid ORFs (Fig. 3A), 2,063 chromosomal ORFs (Fig. 3B), or 40 plasmid ORFs (Fig. 3C). This analysis confirmed the RFLP-based genomic groupings (I to VI) as previously proposed by Hendrix et al. (33). Although, the Id and He isolates both harbor QpH1, they were more related to each other than to genomic group I isolates that also carry this plasmid. Indeed, if only chromosomal polymorphisms are considered, the Du 7E9-12 isolate was most closely related to genomic group I; however, when considering just plasmid polymorphisms, this isolate was most similar to genomic group IV isolates. Genomic group V isolates, having the greatest number of polymorphic ORFs with 87, were most divergent from genomic group I. Q321 and Le B isolates were most closely related to genomic group IV isolates; however, unique polymorphisms indicate that they represent two novel

mosomal ORFs (with 93 of 102 variable characters being parsimony informative and equally weighted) (B), or plasmid ORFs (with all 30 variable characters being parsimony informative and equally weighted) (C). One of the two most parsimonious trees for each of these groups is depicted. (See Fig. S2 in the supplemental material for the alternative tree.) Proposed genomic groups (I to VIII) are shown in panel A. The horizontal bar indicates the number of character changes.

genomic groups of *C. burnetii* referred to here as genomic groups VII and VIII, respectively.

DISCUSSION

In this study we used CGH to detail the extent of genomic diversity of 24 *C. burnetii* isolates obtained from disparate geographical and environmental sources. In all isolates combined, 139 ORFs are polymorphic relative to the NMI reference isolate, with the majority being complete ORF deletions. Twenty-five polymorphisms are partial ORF deletions, point mutations, or small insertions within ORFs that also result in no microarray hybridization signal, with a small number of these predicted to be functional. The lack of broad *C. burnetii* genetic variance is in keeping with the organism's obligate intracellular lifestyle that limits opportunities for genetic exchange. Indeed, as previously described (63), *C. burnetii* ORFs involved in a competence system for the uptake of DNA, i.e., ComA (CBU0855), ComF (CBU0464), and ComE (CBU0532), contain frameshifts, suggesting the organism is no longer naturally competent. In contrast, *Legionella pneumophila*, a phylogenetically close relative of *C. burnetii*, is naturally competent (69).

Consistent with other obligate intracellular bacteria (3, 23, 55, 84), and as proposed by Seshadri et al. (63), our results suggest that the *C. burnetii* genome is undergoing reductive evolution. Accordingly, we found very few polymorphisms resulting from DNA insertion. The *C. burnetii* genome contains 83 pseudogenes, 79 of which are represented on the RMLchip a. Twelve pseudogenes (15.2%) are completely or partially deleted in at least one *C. burnetii* isolate. Conversely, only 6.9% of coding ORFs are deleted; thus, nonfunctional ORFs are apparently being targeted for deletion at a higher frequency than coding ORFs. Moreover, among deleted functional ORFs, there appears to be a bias toward deletion of those with potential functional redundancy. For example, two of five predicted nucleotidyltransferases are missing in Q321 and Le B isolates. Other examples include the deletion of 4 of 14 ankyrin repeat proteins, 2 of 21 predicted ATPases, and 2 of 5 multi-drug resistance B proteins.

Hendrix et al. (33) were equivocal on the genetic distance between genomic groups I, II, and III as each had a distinct RFLP pattern, but all harbored the same plasmid (QpH1). The ability of CGH to screen every ORF in the *C. burnetii* genome allowed us to confirm that isolates within genomic groups I, II, and III are indeed genetically distinct. Eight and four chromosomal polymorphisms distinguish genomic group II and III isolates, respectively, from genomic group I isolates, and six polymorphisms distinguish genomic group II isolates from genomic group III isolates. Phylogenetic trees constructed using CGH data suggest a divergent evolution whereby genomic group I is the ancestor of genomic group III and genomic group III is the ancestor of genomic group II (Fig. 3). Genomic group I isolates were acquired from geographically diverse locations including Australia, United States, Africa, Turkey, and Panama, with the dates of isolation ranging from 1935 to 1967. Interestingly, our results indicate that the gene content of these isolates is exactly the same relative to NMI, suggesting a worldwide spread of a common ancestor that has since undergone limited evolution. *C. burnetii* genome content can also vary among iso-

lates obtained at a similar time and location. Specifically, Du 7E9-12 (rodent isolate) and Du 5G61-63 (tick isolate) were both obtained in Dugway, Utah, in 1958, but only Du 7E9-12 has polymorphisms relative to NMI that are largely the consequence of its carriage of the plasmid QpDG. We also determined the evolutionary relationships of four isolates previously ungrouped by RFLP analysis. BDT and Du 5G61-63 have no ORF polymorphisms relative to genomic group I isolates, and hence, fall within this genomic group. Q321 and Le B are highly similar to the genomic group IV isolates, suggesting a recent common ancestor, yet they have diverged enough to form two new genomic groups (VII and VIII, respectively). Interestingly, the only difference between Le B and Q321 is polymorphism 2, which is present in Le B and genomic group V isolates. Sequencing indicates this polymorphism resulted from insertion of an IS1111 transposon-like sequence at the same location. Genomic groups IV, V, VII, and VIII comprise a distinct clade from genomic groups I, II, III and VI. Within this clade, genomic group V is the most divergent from NMI, being most closely related to the genomic groups IV, VII, and VIII.

Although predisposing host factors are clearly important in the development of human acute or chronic Q fever (32, 51), *C. burnetii* isolates also exhibit distinct phenotypes in terms of infectivity of cell culture (53), cytopathic effects on host cells (53, 66), and pathogenicity in animal models of Q fever (43, 64, 68). The general consensus of these studies is that the NMI reference isolate (genomic group I), biologically representative of human acute disease isolates, is more infectious, grows faster in cell culture, and is more virulent in animal models of infection than chronic disease isolates of genomic groups IV and V. Evidence of *C. burnetii* pathotype-specific ORFs was recently provided by Glazunova et al. (24), who found by multi-spacer sequence typing analysis that both plasmid type and genotype correlate with disease outcome. The CGH identification in the present study of polymorphic ORFs associated with chronic disease isolates and isolates of attenuated virulence allows speculation concerning their roles in *Coxiella* pathogenicity and disease outcome.

The degree of variance between NMI and chronic disease isolates of genomic groups IV and V ranged from 2.3% (51 polymorphic ORFs) to 4.1% (87 polymorphic ORFs), respectively, with most polymorphic ORFs encoding hypothetical proteins. Only seven chromosomal and nine plasmid ORFs are polymorphic in both genomic groups, suggesting the loss of these ORFs may result in the unique phenotypes of these isolates. The NMI genome contains 14 ANK-encoding ORFs (63) and, although these proteins are rare in prokaryotes, ANKs are also found in the intracellular bacteria *Legionella pneumophila* (9), *Anaplasma* spp. (7, 49), *Ehrlichia ruminantium* (15), *Wolbachia pipientis* (20, 80), and *Rickettsia* spp. (4, 47), where they are suspected of modulating host functions. Four ANK-encoding ORFs are missing in genomic group IV and V isolates, with one (CBU1609) having homology to AnkA of *A. phagocytophilum* (49). In *A. phagocytophilum*-infected human promyelocytic leukemia HL-60 cells, AnkA is secreted, binds host DNA, and possibly regulates transcription (8, 49). The absence of CBU1609 and other ANK-encoding genes in *C. burnetii* chronic disease isolates indicates they are not required for potential modulation of host gene function during

the disease process, with the caveat that ANK proteins may be functionally redundant. Two ORFs (CBU1804 and CBU1805) encoding proteins with similarity to the quorum-sensing regulator LuxR were deleted in genomic group V isolates; however, the lack of additional ORFs related to quorum sensing within the *C. burnetii* genome suggests these ORFs serve other regulatory functions. Although one or a few polymorphic ORFs may explain the association of *C. burnetii* genomic groups with chronic disease, another possibility is that deleted ORFs in toto simply result in the slower growth of these isolates and, consequently, the stimulation of an attenuated immune response. Indeed, slow growth resulting from a dramatically reduced genome is thought to contribute to *Mycobacterium leprae*'s known ability to cause chronic infection (13, 81).

Du 7E9-12 of genomic group VI represents a naturally occurring isolate that is weakly pathogenic in a guinea pig model of infection (68). It contains six chromosomal and eight plasmid polymorphisms relative to NMI. However, only one unique polymorphism is predicted to result in loss of a functional ORF. It consists of a partial deletion of CBU0478 which encodes a hypothetical protein. This presents the possibility that CBU0478 is necessary for virulence. There is precedence for small genetic differences conferring dramatically different virulence properties in obligate intracellular bacterial pathogens. For example, the gene contents of *Chlamydia trachomatis* and *C. muridarum* (formally *C. trachomatis*, biovar mouse pneumonitis), human and mouse specific pathogens, respectively, are >99% identical (52). However, these pathogens have a strict tropism for their respective animal hosts, a phenomenon apparently conferred by just a few genes in a small chlamydial plasticity zone (45).

Collectively, 10 ORFs are polymorphic in acute disease isolates of genomic groups II and III. Three ORFs—CBU1991, CBU1992, and CBU0598—predicted to encode RelB, RelE, and a MutT/NUDIX family protein, respectively, are polymorphic in the He isolate of genomic group II. The loss of RelB and RelE, which regulate the stringent response (11, 12), might affect the stationary-phase physiology of the He isolate and, consequently, development of the stable small cell developmental form (14). The absence of CBU0598, might limit the He isolate's ability to detoxify hazardous materials or prevent unbalanced buildup of normal metabolites (6), thereby detrimentally affecting its growth. Both hypotheses should be testable using in vitro infection models.

We recognize caveats in our attempt to correlate *C. burnetii* genetic polymorphisms with specific biological and virulence properties of isolates. First and foremost, our analysis is unidirectional in that all isolates were compared to the sequenced NMI isolate. As such, we cannot ascertain the level of novel coding capacity carried by tested isolates, a distinct possibility considering the range of isolate genome sizes roughly determined by pulse gel electrophoresis (78). Second, the tiled oligonucleotide probe sets on the RMLchip cover ca. 30% of the NMI coding capacity. Therefore, mutations or deletions outside of this region will not be detected. It is also possible that some insertions and point mutations within tiled regions were not detected. We were initially concerned that genomic nucleotide sequence variation between NMI and test isolates might result in negative hybridization signals for intact and functional genes. However, nearly all polymorphisms were

confirmed by PCR validation as true deletions, with only the occasional polymorphism resulting from synonymous sequence variation.

Although the impact of most polymorphic ORFs on *C. burnetii* virulence is speculative, deletion of ORFs involved in LPS biosynthesis is clearly associated with attenuated virulence (51). Our CGH results for NMC and NMII producing truncated LPS are consistent with previous studies showing each variant with a single large chromosomal deletion (35, 76). The NMC deletion includes the 21 ORFs missing in NMII and extends by two ORFs on each side of the NMII deletion (35). Most deleted ORFs are predicted to function in O-antigen biosynthesis, and their absence explains the intermediate truncation of NMC LPS. However, their absence does not explain why NMII, with its smaller deletion, produces a more truncated LPS than NMC. NMII likely has an additional point/frame shift mutation, small deletion, or transposon insertion in a gene early in the LPS biosynthetic pathway that is not detected by CGH. An analogous situation may exist with the Au isolate, which produces a severely truncated LPS similar to NMII but is identical to NMI by CGH. Thus, our data confirm not only that large chromosomal deletions are not required for *C. burnetii* LPS phase variation, as previously proposed by Thompson et al. (72), but also that the avirulence of NMII *C. burnetii* appears to be entirely attributable to deficient LPS production since other ORF deletions were not detected.

Plasmid polymorphisms specific to QpH1, QpRS, QpDV, and integrated plasmid-like sequences were confirmed via comparison with their deciphered sequences (38, 63, 79), while polymorphisms of QpDG are described here for the first time. Nine predicted functional plasmid ORFs are conserved among all plasmid types or integrated-plasmid sequences. Interestingly, two of these ORFs, CBUA0008 and CBUA0010, encode proteins that exhibit similarity to phage proteins involved in tail assembly and site-specific recombination, respectively. These ORFs may be remnants of a phage that had integrated into the *C. burnetii* plasmid. Flanked by these phage genes is a degenerate ORF (CBUA0009) annotated as a putative toxin. This protein has the highest similarity to insecticidal and nematocidal toxin synthesized by the waterborne bacterium *Chromobacterium violaceum* (16) and nematode symbionts *Xenorhabdus bovienii* and *Photorhabdus luminescens* (10, 19), suggesting it was acquired by *C. burnetii* via horizontal transfer mediated by phage integration. This ORF may have been beneficial to a *C. burnetii* progenitor with an insect host range. Interestingly, the closest relative of *C. burnetii* is *Rickettsiella grylli*, a cricket pathogen (54). Excluding phage-related ORFs, seven plasmid ORFs are conserved among all *C. burnetii* isolates examined in the present study, suggesting a critical role in some aspect of *C. burnetii* intracellular parasitism. Only CBUA0011 has an annotated function, with similarity to *ripX* of *Bacillus subtilis*, a gene involved in chromosomal partitioning (61).

Movement of mobile genetic elements clearly contributes to *C. burnetii* genomic plasticity. *C. burnetii* contains a large number of insertion sequences (IS elements) (63) relative to most obligate intracellular bacteria (13, 47, 67, 70). Of the 51 single-event polymorphisms, 9 are within five ORFs of a transposase or contain a transposase, indicating that these IS-type elements have a role in generating *C. burnetii* genetic diversity. Indeed, Southern blotting shows a range in the number of IS elements

between *C. burnetii* isolate genomic groups (manuscript in preparation). By this analysis, isolates within a genomic group have identical hybridization patterns, supporting the phylogenetic relationships revealed by CGH. Movement and/or loss of these mobile elements could be a mechanism by which *C. burnetii* rearranges and/or reduces its genome to better suit its intracellular lifestyle.

In summary, the present study provides the first comprehensive whole-genome genotyping of *C. burnetii*. We show that, in general, the genomes of *C. burnetii* isolates obtained from a wide range of biologically and geographically diverse sources are highly conserved. However, notable ORF polymorphisms occur that may contribute to the virulence potential and other biological properties of *C. burnetii* isolates. CGH assessment of *C. burnetii* genome content can help to identify cross-protective subunit vaccine candidates for protection against Q fever and facilitate the development of new diagnostic tools and, as an epidemiological method, to classify clinical, forensic, and environmental isolates. Indeed, in conjunction with whole-genome amplification, rapid whole-genome typing of isolates in clinical samples can be achieved without propagation of the organism. Finally, a CGH approach to identify potential pathotype-specific virulence genes is especially valuable considering the lack of genetic systems for *C. burnetii*.

ACKNOWLEDGMENTS

We thank Ted Hackstadt and John H. Carlson for critical reading of the manuscript and Anita Mora and Gary Hettrick for graphic illustrations.

This research was supported by the Intramural Research Program of the National Institutes of Health, National Institute of Allergy and Infectious Diseases.

REFERENCES

- Amano, K., and J. C. Williams. 1984. Chemical and immunological characterization of lipopolysaccharides from phase I and phase II *Coxiella burnetii*. *J. Bacteriol.* **160**:994–1002.
- Amano, K., J. C. Williams, S. R. Missler, and V. N. Reinhold. 1987. Structure and biological relationships of *Coxiella burnetii* lipopolysaccharides. *J. Biol. Chem.* **262**:4740–4747.
- Andersson, J. O., and S. G. Andersson. 2001. Pseudogenes, junk DNA, and the dynamics of *Rickettsia* genomes. *Mol. Biol. Evol.* **18**:829–839.
- Andersson, S. G., A. Zomorodipour, J. O. Andersson, T. Sicheritz-Ponten, U. C. Alsmark, R. M. Podowski, A. K. Naslund, A. S. Eriksson, H. H. Winkler, and C. G. Kurland. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**:133–140.
- Andoh, M., H. Nagaoka, T. Yamaguchi, H. Fukushima, and K. Hirai. 2004. Comparison of Japanese isolates of *Coxiella burnetii* by PCR-RFLP and sequence analysis. *Microbiol. Immunol.* **48**:971–975.
- Bessman, M. J., D. N. Frick, and S. F. O'Handley. 1996. The MutT proteins or "Nudix" hydrolases, a family of versatile, widely distributed, "housecleaning" enzymes. *J. Biol. Chem.* **271**:25059–25062.
- Brayton, K. A., L. S. Kappmeyer, D. R. Herndon, M. J. Dark, D. L. Tibbals, G. H. Palmer, T. C. McGuire, and D. P. Knowles, Jr. 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc. Natl. Acad. Sci. USA* **102**:844–849.
- Caturegli, P., K. M. Asanovich, J. J. Walls, J. S. Bakken, J. E. Madigan, V. L. Popov, and J. S. Dumler. 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. *Infect. Immun.* **68**:5277–5283.
- Cazalet, C., C. Rusniok, H. Bruggemann, N. Zidane, A. Magnier, L. Ma, M. Tichit, S. Jarraud, C. Bouchier, F. Vandenesch, F. Kunst, J. Etienne, P. Glaser, and C. Buchrieser. 2004. Evidence in the *Legionella pneumophila* genome for exploitation of host cell functions and high genome plasticity. *Nat. Genet.* **36**:1165–1173.
- Chen, G., Y. Zhang, J. Li, G. B. Dunphy, Z. K. Punja, and J. M. Webster. 1996. Chitinase activity of *Xenorhabdus* and *Photorhabdus* species, bacterial associates of entomopathogenic nematodes. *J. Invertebr. Pathol.* **68**:101–108.
- Christensen, S. K., and K. Gerdes. 2004. Delayed-relaxed response explained by hyperactivation of RelE. *Mol. Microbiol.* **53**:587–597.
- Christensen, S. K., M. Mikkelsen, K. Pedersen, and K. Gerdes. 2001. RelE, a global inhibitor of translation, is activated during nutritional stress. *Proc. Natl. Acad. Sci. USA* **98**:14328–14333.
- Cole, S. T., K. Eglmeier, J. Parkhill, K. D. James, N. R. Thomson, P. R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M. A. Quail, M. A. Rajandream, K. M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J. R. Woodward, and B. G. Barrell. 2001. Massive gene decay in the leprosy bacillus. *Nature* **409**:1007–1011.
- Coleman, S. A., E. R. Fischer, D. Howe, D. J. Mead, and R. A. Heinzen. 2004. Temporal analysis of *Coxiella burnetii* morphological differentiation. *J. Bacteriol.* **186**:7344–7352.
- Collins, N. E., J. Liebenberg, E. P. de Villiers, K. A. Brayton, E. Louw, A. Pretorius, F. E. Faber, H. van Heerden, A. Josemans, M. van Kleef, H. C. Steyn, M. F. van Strijp, E. Zweygarth, F. Jongejans, J. C. Maillard, D. Berthier, M. Botha, F. Joubert, C. H. Corton, N. R. Thomson, M. T. Allsopp, and B. A. Allsopp. 2005. The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number. *Proc. Natl. Acad. Sci. USA* **102**:838–843.
- Consortium, B. N. G. P. 2003. The complete genome sequence of *Chromobacterium violaceum* reveals remarkable and exploitable bacterial adaptability. *Proc. Natl. Acad. Sci. USA* **100**:11660–11665.
- Davis, G. E., and H. R. Cox. 1938. A filter-passing infectious agent isolated from ticks. I. Isolation from *Dermacentor andersoni*, reactions with animals, and filtration experiments. *Public Health Rep.* **53**:2259–2276.
- Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren. 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. *Genome Res.* **11**:1706–1715.
- Duchaude, E., C. Rusniok, L. Frangeul, C. Buchrieser, A. Givaudan, S. Taourit, S. Bocs, C. Boursaux-Eude, M. Chandler, J. F. Charles, E. Dassa, R. Derose, S. Derzelle, G. Freyssinet, S. Gaudriault, C. Medigue, A. Lanois, K. Powell, P. Siguier, R. Vincent, V. Wingate, M. Zouine, P. Glaser, N. Boemare, A. Danchin, and F. Kunst. 2003. The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*. *Nat. Biotechnol.* **21**:1307–1313.
- Foster, J., M. Ganatra, I. Kamal, J. Ware, K. Makarova, N. Ivanova, A. Bhattacharyya, V. Kapatral, S. Kumar, J. Posfai, T. Vincze, J. Ingram, L. Moran, A. Lapidus, M. Omelchenko, N. Kyrpides, E. Ghedin, S. Wang, E. Goldsman, V. Joukov, O. Ostrovskaya, K. Tsukerman, M. Mazur, D. Comb, E. Koonin, and B. Slatko. 2005. The *Wolbachia* genome of *Brugia malayi*: endosymbiont evolution within a human pathogenic nematode. *PLoS Biol.* **3**:e121.
- Fournier, P. E., T. J. Marrie, and D. Raoult. 1998. Diagnosis of Q fever. *J. Clin. Microbiol.* **36**:1823–1834.
- Fukuya, S., H. Mizoguchi, T. Tobe, and H. Mori. 2004. Extensive genomic diversity in pathogenic *Escherichia coli* and *Shigella* strains revealed by comparative genomic hybridization microarray. *J. Bacteriol.* **186**:3911–3921.
- Gil, R., B. Sabater-Munoz, A. Latorre, F. J. Silva, and A. Moya. 2002. Extreme genome reduction in *Buchnera* spp.: toward the minimal genome needed for symbiotic life. *Proc. Natl. Acad. Sci. USA* **99**:4454–4458.
- Glazunova, O., V. Roux, O. Freylikman, Z. Sekeyova, G. Fournous, J. Tycka, N. Tokarevich, E. Kovacova, T. J. Marrie, and D. Raoult. 2005. *Coxiella burnetii* genotyping. *Emerg. Infect. Dis.* **11**:1211–1217.
- Goebel, M., and M. Yanagida. 1991. The TPR snap helix: a novel protein repeat motif from mitosis to transcription. *Trends Biochem. Sci.* **16**:173–177.
- Gray, K. M. 1997. Intercellular communication and group behavior in bacteria. *Trends Microbiol.* **5**:184–188.
- Hackstadt, T. 1986. Antigenic variation in the phase I lipopolysaccharide of *Coxiella burnetii* isolates. *Infect. Immun.* **52**:337–340.
- Hackstadt, T. 1990. The role of lipopolysaccharides in the virulence of *Coxiella burnetii*. *Ann. N. Y. Acad. Sci.* **590**:27–32.
- Hackstadt, T., R. Messer, W. Cieplak, and M. G. Peacock. 1992. Evidence for proteolytic cleavage of the 120-kilodalton outer membrane protein of rickettsiae: identification of an avirulent mutant deficient in processing. *Infect. Immun.* **60**:159–165.
- Hackstadt, T., M. G. Peacock, P. J. Hitchcock, and R. L. Cole. 1985. Lipopolysaccharide variation in *Coxiella burnetii*: intrastain heterogeneity in structure and antigenicity. *Infect. Immun.* **48**:359–365.
- Heinzen, R., G. L. Stiegler, L. L. Whiting, S. A. Schmitt, L. P. Mallavia, and M. E. Frazier. 1990. Use of pulsed field gel electrophoresis to differentiate *Coxiella burnetii* strains. *Ann. N. Y. Acad. Sci.* **590**:504–513.
- Helbig, K. J., S. L. Heatley, R. J. Harris, C. G. Mullighan, P. G. Bardy, and B. P. Marmion. 2003. Variation in immune response genes and chronic Q fever. Concepts: preliminary test with post-Q fever fatigue syndrome. *Genes Immun.* **4**:82–85.

33. Hendrix, L. R., J. E. Samuel, and L. P. Mallavia. 1991. Differentiation of *Coxiella burnetii* isolates by analysis of restriction-endonuclease-digested DNA separated by SDS-PAGE. *J. Gen. Microbiol.* **137**:269–276.
34. Hendrix, L. R., J. E. Samuel, and L. P. Mallavia. 1990. Identification and cloning of a 27-kDa *Coxiella burnetii* immunoreactive protein. *Ann. N. Y. Acad. Sci.* **590**:534–540.
35. Hoover, T. A., D. W. Culp, M. H. Vodkin, J. C. Williams, and H. A. Thompson. 2002. Chromosomal DNA deletions explain phenotypic characteristics of two antigenic variants, phase II and RSA 514 (crazy), of the *Coxiella burnetii* nine mile strain. *Infect. Immun.* **70**:6726–6733.
36. Jager, C., H. Willems, D. Thiele, and G. Baljer. 1998. Molecular characterization of *Coxiella burnetii* isolates. *Epidemiol. Infect.* **120**:157–164.
37. Joyce, E. A., K. Chan, N. R. Salama, and S. Falkow. 2002. Redefining bacterial populations: a post-genomic reformation. *Nat. Rev. Genet.* **3**:462–473.
38. Lautenschlager, S., H. Willems, C. Jager, and G. Baljer. 2000. Sequencing and characterization of the cryptic plasmid QpRS from *Coxiella burnetii*. *Plasmid* **44**:85–88.
39. Lipshutz, R. J., S. P. Fodor, T. R. Gingeras, and D. J. Lockhart. 1999. High density synthetic oligonucleotide arrays. *Nat. Genet.* **21**:20–24.
40. Mallavia, L. P. 1991. Genetics of rickettsiae. *Eur. J. Epidemiol.* **7**:213–221.
41. Maurin, M., and D. Raoult. 1999. Q fever. *Clin. Microbiol. Rev.* **12**:518–553.
42. Minnick, M. F., C. L. Small, M. E. Frazier, and L. P. Mallavia. 1991. Analysis of the *cbhE* plasmid gene from acute disease-causing isolates of *Coxiella burnetii*. *Gene* **103**:113–118.
43. Moos, A., and T. Hackstadt. 1987. Comparative virulence of intra- and interstrain lipopolysaccharide variants of *Coxiella burnetii* in the guinea pig model. *Infect. Immun.* **55**:1144–1150.
44. Mosavi, L. K., T. J. Cammett, D. C. Desrosiers, and Z. Y. Peng. 2004. The ankyrin repeat as molecular architecture for protein recognition. *Protein Sci.* **13**:1435–1448.
45. Nelson, D. E., D. P. Virok, H. Wood, C. Roshick, R. M. Johnson, W. M. Whitmire, D. D. Crane, O. Steele-Mortimer, L. Laszlo Kari, G. McClarty, and H. D. Caldwell. 2005. Chlamydial IFN- γ immune evasion is linked to host infection tropism. *Proc. Natl. Acad. Sci. USA* **102**:10658–10663.
46. Nguyen, S. V., and K. Hirai. 1999. Differentiation of *Coxiella burnetii* isolates by sequence determination and PCR-restriction fragment length polymorphism analysis of isocitrate dehydrogenase gene. *FEMS Microbiol. Lett.* **180**:249–254.
47. Ogata, H., S. Audic, P. Renesto-Audiffren, P. E. Fournier, V. Barbe, D. Samson, V. Roux, P. Cossart, J. Weissenbach, J. M. Claverie, and D. Raoult. 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* **293**:2093–2098.
48. Ong, C., C. H. Ooi, D. Wang, H. Chong, K. C. Ng, F. Rodrigues, M. A. Lee, and P. Tan. 2004. Patterns of large-scale genomic variation in virulent and avirulent *Burkholderia* species. *Genome Res.* **14**:2295–2307.
49. Park, J., K. J. Kim, K. S. Choi, D. J. Grab, and J. S. Dumler. 2004. *Anaplasma phagocytophilum* Anka binds to granulocyte DNA and nuclear proteins. *Cell. Microbiol.* **6**:743–751.
50. Porwollik, S., R. M. Wong, and M. McClelland. 2002. Evolutionary genomics of *Salmonella*: gene acquisitions revealed by microarray analysis. *Proc. Natl. Acad. Sci. USA* **99**:8956–8961.
51. Raoult, D., T. Marrie, and J. Mege. 2005. Natural history and pathophysiology of Q fever. *Lancet Infect. Dis.* **5**:219–226.
52. Read, T. D., R. C. Brunham, C. Shen, S. R. Gill, J. F. Heidelberg, O. White, E. K. Hickey, J. Peterson, T. Utterback, K. Berry, S. Bass, K. Linher, J. Weidman, H. Khouri, B. Craven, C. Bowman, R. Dodson, M. Gwinn, W. Nelson, R. DeBoy, J. Kolonay, G. McClarty, S. L. Salzberg, J. Eisen, and C. M. Fraser. 2000. Genome sequences of *Chlamydia trachomatis* MoPn and *Chlamydia pneumoniae* AR39. *Nucleic Acids Res.* **28**:1397–1406.
53. Roman, M. J., H. A. Crissman, W. A. Samsonoff, K. E. Hechemy, and O. G. Baca. 1991. Analysis of *Coxiella burnetii* isolates in cell culture and the expression of parasite-specific antigens on the host membrane surface. *Acta Virol.* **35**:503–510.
54. Roux, V., M. Bergoin, N. Lamaze, and D. Raoult. 1997. Reassessment of the taxonomic position of *Rickettsia grylli*. *Int. J. Syst. Bacteriol.* **47**:1255–1257.
55. Sakharkar, K. R., P. K. Dhar, and V. T. Chow. 2004. Genome reduction in prokaryotic obligatory intracellular parasites of humans: a comparative analysis. *Int. J. Syst. Evol. Microbiol.* **54**:1937–1941.
56. Salama, N., K. Guillemin, T. K. McDaniel, G. Sherlock, L. Tompkins, and S. Falkow. 2000. A whole-genome microarray reveals genetic diversity among *Helicobacter pylori* strains. *Proc. Natl. Acad. Sci. USA* **97**:14668–14673.
57. Sallstrom, B., and S. G. Andersson. 2005. Genome reduction in the alpha-Proteobacteria. *Curr. Opin. Microbiol.* **8**:1–7.
58. Samuel, J. E., M. E. Frazier, M. L. Kahn, L. S. Thomashow, and L. P. Mallavia. 1983. Isolation and characterization of a plasmid from phase I *Coxiella burnetii*. *Infect. Immun.* **41**:488–493.
59. Samuel, J. E., M. E. Frazier, and L. P. Mallavia. 1985. Correlation of plasmid type and disease caused by *Coxiella burnetii*. *Infect. Immun.* **49**:775–779.
60. Savinelli, E. A., and L. P. Mallavia. 1990. Comparison of *Coxiella burnetii* plasmids to homologous chromosomal sequences present in a plasmidless endocarditis-causing isolate. *Ann. N. Y. Acad. Sci.* **590**:523–533.
61. Sciochetti, S. A., P. J. Piggot, D. J. Sherratt, and G. Blakely. 1999. The *ripX* locus of *Bacillus subtilis* encodes a site-specific recombinase involved in proper chromosome partitioning. *J. Bacteriol.* **181**:6053–6062.
62. Sekeyova, Z., V. Roux, and D. Raoult. 1999. Intraspecies diversity of *Coxiella burnetii* as revealed by *comI* and *mucZ* sequence comparison. *FEMS Microbiol. Lett.* **180**:61–67.
63. Seshadri, R., I. T. Paulsen, J. A. Eisen, T. D. Read, K. E. Nelson, W. C. Nelson, N. L. Ward, H. Tettelin, T. M. Davidsen, M. J. Beanan, R. T. Deboy, S. C. Daugherty, L. M. Brinkac, R. Madupu, R. J. Dodson, H. M. Khouri, K. H. Lee, H. A. Carty, D. Scanlan, R. A. Heinzen, H. A. Thompson, J. E. Samuel, C. M. Fraser, and J. F. Heidelberg. 2003. Complete genome sequence of the Q-fever pathogen *Coxiella burnetii*. *Proc. Natl. Acad. Sci. USA* **100**:5455–5460.
64. Stein, A., C. Louveau, H. Lepidi, F. Ricci, P. Baylac, B. Davoust, and D. Raoult. 2005. Q fever pneumonia: virulence of *Coxiella burnetii* pathovars in a murine model of aerosol infection. *Infect. Immun.* **73**:2469–2477.
65. Stein, A., and D. Raoult. 1993. Lack of pathotype specific gene in human *Coxiella burnetii* isolates. *Microb. Pathog.* **15**:177–185.
66. Stein, A., and D. Raoult. 1992. Phenotypic and genotypic heterogeneity of eight new human *Coxiella burnetii* isolates. *Acta Virol.* **36**:7–12.
67. Stephens, R. S., S. Kalman, C. Lammel, J. Fan, R. Marathe, L. Aravind, W. Mitchell, L. Olinger, R. L. Tatusov, Q. Zhao, E. V. Koonin, and R. W. Davis. 1998. Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**:754–759.
68. Stoenner, H. G., and D. B. Lackman. 1960. The biologic properties of *Coxiella burnetii* isolated from rodents collected in Utah. *Am. J. Hyg.* **71**:45–51.
69. Stone, B. J., and Y. A. Kwaik. 1999. Natural competence for DNA transformation by *Legionella pneumophila* and its association with expression of type IV pili. *J. Bacteriol.* **181**:1395–1402.
70. Tamas, L., L. Klasson, B. Canback, A. K. Naslund, A. S. Eriksson, J. J. Wernegreen, J. P. Sandstrom, N. A. Moran, and S. G. Andersson. 2002. 50 million years of genomic stasis in endosymbiotic bacteria. *Science* **296**:2376–2379.
71. Thiele, D., and H. Willems. 1994. Is plasmid based differentiation of *Coxiella burnetii* in 'acute' and 'chronic' isolates still valid? *Eur. J. Epidemiol.* **10**:427–434.
72. Thompson, H. A., T. A. Hoover, M. H. Vodkin, and E. I. Shaw. 2003. Do chromosomal deletions in the lipopolysaccharide biosynthetic regions explain all cases of phase variation in *Coxiella burnetii* strains? An update. *Ann. N. Y. Acad. Sci.* **990**:664–670.
73. To, H., A. Hotta, T. Yamaguchi, H. Fukushima, and K. Hirai. 1998. Antigenic characteristic of the lipopolysaccharides of *Coxiella burnetii* isolates. *J. Vet. Med. Sci.* **60**:267–270.
74. Toman, R., and L. Skultety. 1996. Structural study on a lipopolysaccharide from *Coxiella burnetii* strain Nine Mile in avirulent phase II. *Carbohydr. Res.* **283**:175–185.
75. Valkova, D., and J. Kazar. 1995. A new plasmid (QpDV) common to *Coxiella burnetii* isolates associated with acute and chronic Q fever. *FEMS Microbiol. Lett.* **125**:275–280.
76. Vodkin, M. H., and J. C. Williams. 1986. Overlapping deletion in two spontaneous phase variants of *Coxiella burnetii*. *J. Gen. Microbiol.* **132**:2587–2594.
77. Vodkin, M. H., J. C. Williams, and E. H. Stephenson. 1986. Genetic heterogeneity among isolates of *Coxiella burnetii*. *J. Gen. Microbiol.* **132**:455–463.
78. Willems, H., D. C. Thiele, M. Burger, W. Ritter, C. Daugherty, and H. Krauss. 1996. Molecular biology of *Coxiella burnetii*, p. 363–378. In J. Kazar and R. Toman (ed.), *Rickettsia and rickettsial diseases*. Slovak Academy of Sciences, Bratislava, Slovakia.
79. Willems, H., M. Ritter, C. Jager, and D. Thiele. 1997. Plasmid-homologous sequences in the chromosome of plasmidless *Coxiella burnetii* Scurry Q217. *J. Bacteriol.* **179**:3293–3297.
80. Wu, M., L. V. Sun, J. Vamathevan, M. Riegler, R. Deboy, J. C. Brownlie, E. A. McGraw, W. Martin, C. Esser, N. Ahmadinejad, C. Wiegand, R. Madupu, M. J. Beanan, L. M. Brinkac, S. C. Daugherty, A. S. Durkin, J. F. Kolonay, W. C. Nelson, Y. Mohamoud, P. Lee, K. Berry, M. B. Young, T. Utterback, J. Weidman, W. C. Nierman, I. T. Paulsen, K. E. Nelson, H. Tettelin, S. L. O'Neill, and J. A. Eisen. 2004. Phylogenomics of the reproductive parasite *Wolbachia pipiensis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.* **2**:e69.
81. Young, D., T. Huseell, and G. Dougan. 2002. Chronic bacterial infections: living with unwanted guests. *Nat. Immunol.* **3**:1026–1032.
82. Zhang, G., H. To, K. E. Russell, L. R. Hendrix, T. Yamaguchi, H. Fukushima, K. Hirai, and J. E. Samuel. 2005. Identification and characterization of an immunodominant 28-kilodalton *Coxiella burnetii* outer membrane protein specific to isolates associated with acute disease. *Infect. Immun.* **73**:1561–1567.
83. Zhang, G. Q., H. To, T. Yamaguchi, H. Fukushima, and K. Hirai. 1997. Differentiation of *Coxiella burnetii* by sequence analysis of the gene (*comI*) encoding a 27-kDa outer membrane protein. *Microbiol. Immunol.* **41**:871–877.
84. Zomorodipour, A., and S. G. Andersson. 1999. Obligate intracellular parasites: *Rickettsia prowazekii* and *Chlamydia trachomatis*. *FEBS Lett.* **452**:11–15.